

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 02:16:01 ; Search time 554 seconds  
(without alignments)  
6979.281 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363

Sequence: 1 tcgagccgccttcacaggga.....taataaagatggttcgattc 2363

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813859 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 2363   | 100.0       | 2363   | 3     | US-09-742-703-3     |
| 2          | 1593.6 | 67.4        | 1783   | 3     | US-09-510-738A-188  |
| 3          | 1593.6 | 67.4        | 1783   | 4     | US-09-861-966-188   |
| 4          | 1593.6 | 67.4        | 1783   | 4     | US-09-915-048-188   |
| 5          | 1593.6 | 67.4        | 1783   | 4     | US-09-949-016-1311  |
| 6          | 1269   | 53.7        | 1615   | 4     | US-09-820-002-1     |
| 7          | 1008.2 | 42.7        | 1605   | 2     | US-09-000-846-1     |
| 8          | 806.8  | 34.1        | 30337  | 4     | US-09-949-016-13053 |
| 9          | 391.2  | 16.6        | 614    | 3     | US-09-280-116-71    |
| 10         | 325    | 13.8        | 21784  | 4     | US-09-820-002-3     |
| 11         | 192    | 8.1         | 601    | 4     | US-09-820-002-15    |
| 12         | 192    | 8.1         | 601    | 4     | US-09-949-016-45925 |
| 13         | 191    | 8.1         | 601    | 4     | US-09-820-002-16    |
| 14         | 169.4  | 7.2         | 601    | 4     | US-09-820-002-14    |
| 15         | 169.4  | 7.2         | 601    | 4     | US-09-949-016-45924 |
| 16         | 148.4  | 6.3         | 1434   | 4     | US-10-177-661-1     |
| 17         | 148.4  | 6.3         | 1434   | 4     | US-09-879-792-11    |
| 18         | 145.8  | 6.2         | 1341   | 4     | US-10-177-661-3     |
| 19         | 141.4  | 6.0         | 2443   | 3     | US-09-518-046-1     |
| 20         | 130    | 5.5         | 2440   | 4     | US-09-949-016-5210  |
| 21         | 130    | 5.5         | 2440   | 4     | US-09-949-016-5211  |
| 22         | 130    | 5.5         | 2440   | 4     | US-09-949-016-5212  |
| 23         | 127.4  | 5.4         | 2479   | 3     | US-09-342-749-29    |
| 24         | 127.4  | 5.4         | 2479   | 3     | US-09-691-840-29    |
| 25         | 127.4  | 5.4         | 2479   | 4     | US-09-685-166A-894  |
| 26         | 127.4  | 5.4         | 2479   | 4     | US-09-679-426-894   |
| 27         | 127.4  | 5.4         | 2479   | 4     | US-09-759-143-894   |

#### ALIGNMENTS

|    |       |     |      |   |                    |                    |
|----|-------|-----|------|---|--------------------|--------------------|
| 28 | 124.8 | 5.3 | 1077 | 3 | US-08-807-151-2    | Sequence 2, App1   |
| 29 | 124.8 | 5.3 | 1077 | 3 | US-09-478-957-2    | Sequence 2, App1   |
| 30 | 124.6 | 5.3 | 2544 | 3 | US-09-518-046-3    | Sequence 3, App1   |
| 31 | 124.2 | 5.3 | 2448 | 4 | US-09-949-016-5203 | Sequence 3203, App |
| 32 | 124.2 | 5.3 | 3245 | 4 | US-09-759-143-929  | Sequence 929, App  |
| 33 | 123.4 | 5.2 | 1479 | 3 | US-09-342-749-1    | Sequence 1, App1   |
| 34 | 123.4 | 5.2 | 1479 | 3 | US-09-691-840-1    | Sequence 910, App1 |
| 35 | 121.8 | 5.2 | 1479 | 4 | US-09-759-143-930  | Sequence 930, App  |
| 36 | 120.6 | 5.1 | 1476 | 4 | US-09-759-143-931  | Sequence 931, App  |
| 37 | 120   | 5.1 | 2416 | 3 | US-09-261-416-1    | Sequence 256, App  |
| 38 | 117.2 | 5.0 | 1100 | 4 | US-09-907-794A-256 | Sequence 256, App  |
| 39 | 117.2 | 5.0 | 1100 | 4 | US-09-905-125A-256 | Sequence 256, App  |
| 40 | 117.2 | 5.0 | 1100 | 4 | US-09-902-775A-256 | Sequence 256, App  |
| 41 | 117.2 | 5.0 | 1100 | 4 | US-09-906-700-256  | Sequence 256, App  |
| 42 | 117.2 | 5.0 | 1100 | 4 | US-09-903-603A-256 | Sequence 256, App  |
| 43 | 117.2 | 5.0 | 1100 | 4 | US-09-904-920A-256 | Sequence 256, App  |
| 44 | 117.2 | 5.0 | 1100 | 4 | US-09-899-064-256  | Sequence 256, App  |
| 45 | 117.2 | 5.0 | 1100 | 4 | US-09-905-381A-256 | Sequence 256, App  |

  

|  |                 |
|--|-----------------|
| RESULT 1   | US-09-742-703-3 |
| Sequence 3, Application US/09742703                            |                 |
| Patent No. 6423543   |                 |
| GENERAL INFORMATION:   |                 |
| APPLICANT: Lex M. Cowsett                                      |                 |
| INVENTOR: Patrick Allen Marcotte                               |                 |
| TITLE OF INVENTION: ANTISENSE MODULATION OF HEPESIN EXPRESSION |                 |
| FILE REFERENCE: RTS-0090                                       |                 |
| CURRENT APPLICATION NUMBER: US/09/742,703                      |                 |
| CURRENT FILING DATE: 2000-12-20                                |                 |
| NUMBER OF SEQ ID NOS: 49                                       |                 |
| SEQ ID NO 3  |                 |
| LENGTH: 2363   |                 |
| TYPE: DNA  |                 |
| ORGANISM: Homo sapiens   |                 |
| FEATURE:   |                 |
| NAME/KEY: CDS  |                 |
| LOCATION: (826)..(2079)  |                 |
| US-09-742-703-3  |                 |

  

|                       |                 |               |           |              |
|-----------------------|-----------------|---------------|-----------|--------------|
| Query Match           | 100.0%;         | Score 2363;   | DB 3;     | Length 2363; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;  |           |              |
| Matches 2363;         | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0;      |

  

|    |     |   |     |
|----|-----|---|-----|
| QY | 1   | TCGAGCCCCGCTTTTCAGAGGACCTTAAGAGGCGCCACAGGTGAGGACCTTGCCCTTAC | 60  |
| DB | 1   | TCGAGCCCCGCTTTTCAGAGGACCTTAAGAGGCGCCACAGGTGAGGACCTTGCCCTTAC | 60  |
| QY | 61  | AGGCCCCCAGCCAGCCGCTTGCTCCAGGCGCCGCTGCGGGGACCACTGCTCC        | 120 |
| DB | 61  | AGGCCCCCAGCCAGCCGCTTGCTCCAGGCGCCGCTGCGGGGACCACTGCTCC        | 120 |
| QY | 121 | TGCCAGGCTTGAGAGTGAAGCCGAGCCGAGCACTTAAGAGGCTTCGCCCTTAC       | 180 |
| DB | 121 | TGCCAGGCTTGAGAGTGAAGCCGAGCCGAGCACTTAAGAGGCTTCGCCCTTAC       | 180 |
| QY | 181 | TGAGCCCCAGGTTAAGAGCAAGGCGCCAGACTCAAGTTCAGCCCTTGAAGCAAGGG    | 240 |
| DB | 181 | TGAGCCCCAGGTTAAGAGCAAGGCGCCAGACTCAAGTTCAGCCCTTGAAGCAAGGG    | 240 |
| QY | 241 | TTCCCTCATCCCCCAAGCCAGCTTAATGCCACCTCTTAATAGAGGGTTCTGGAGAC    | 300 |
| DB | 241 | TTCCCTCATCCCCCAAGCCAGCTTAATGCCACCTCTTAATAGAGGGTTCTGGAGAC    | 300 |
| QY | 301 | TGAAGAGGGGCACTATGAGCTTCCCAAGCACTTAAGTGTCTGCTGCTTCTT         | 360 |
| DB | 301 | TGAAGAGGGGCACTATGAGCTTCCCAAGCACTTAAGTGTCTGCTGCTTCTT         | 360 |
| QY | 361 | CAGACTCAGCGCTTGAAGCCCAAGTCTTCTCCCAAGCAAGGAGTTCCAGCCCTCAG    | 420 |

Db 361 CAGACTAGCGCTGGAGCCCAAGTCTTCTCCCAAGACCAAGAGTCCAGCCCTAG 420  
 Qy 421 GCGCTCTCTCTCTATCTAGAGTCTGGCCCCCAATCTCTCTTCCCAAGACTTA 480  
 Db 421 GCGCTCTCTCTCTATCTAGAGTCTGGCCCCCAATCTCTCTTCCCAAGACTTA 480  
 Qy 481 TGATTTAGAGTCTCAGCTGTCTCTCTCCCAACCGGGAGTCTCAGTCCCTCTGAC 540  
 Db 481 TGATTTAGAGTCTCAGCTGTCTCTCTCCCAACCGGGAGTCTCAGTCCCTCTGAC 540  
 Qy 541 CAGGCTCAGGAGTGGGGTCCCAATCCCTGCAATCCAGAGTCCCGGCTGCTGCA 600  
 Db 541 CAGGCTCAGGAGTGGGGTCCCAATCCCTGCAATCCAGAGTCCCGGCTGCTGCA 600  
 Qy 601 GACACTAGACCCATCTTGAACCCAGCCCAATCTGCTGCTGATTCAGGGGTCTCTG 660  
 Db 601 GACACTAGACCCATCTTGAACCCAGCCCAATCTGCTGCTGATTCAGGGGTCTCTG 660  
 Qy 661 CCAAGGCGCAGTCCCTCAGCTGCTGCTGATGAGAGGCTGGAGCTGGGGGCGCAGGACT 720  
 Db 661 CCAAGGCGCAGTCCCTCAGCTGCTGCTGATGAGAGGCTGGAGCTGGGGGCGCAGGACT 720  
 Qy 721 GGGCTGGGCTGGGCTCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
 Db 721 GGGCTGGGCTGGGCTCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
 Qy 781 TGCGCCAGAGAGTCAAGCCAGGAATCAATTAACAAGAGCAGTGAATGCGCCCAAGAGAG 840  
 Db 781 TGCGCCAGAGAGTCAAGCCAGGAATCAATTAACAAGAGCAGTGAATGCGCCCAAGAGAG 840  
 Qy 841 GGTGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 Db 841 GGTGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 Qy 901 CTGCTACTTCTGACAGCCATGCGGGCGGCAATCTGCGGCAATTTGGCTTCTCTCAGG 960  
 Db 901 CTGCTACTTCTGACAGCCATGCGGGCGGCAATCTGCGGCAATTTGGCTTCTCTCAGG 960  
 Qy 961 AGTACCAAGAGAGCGGCTGTATCCCAAGTGAAGTCACTGCGGAGCGTCTGCTGCTGCTG 1020  
 Db 961 AGTACCAAGAGAGCGGCTGTATCCCAAGTGAAGTCACTGCGGAGCGTCTGCTGCTGCTG 1020  
 Qy 1021 TTTTACAAGAGAGAGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
 Db 1021 TTTTACAAGAGAGAGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
 Qy 1081 GCGGAGTCAAGTCTGAGAGAGTGGCTTCTGAGGAGTGAACCTGCAAGCTGAGC 1140  
 Db 1081 GCGGAGTCAAGTCTGAGAGAGTGGCTTCTGAGGAGTGAACCTGCAAGCTGAGC 1140  
 Qy 1141 GTGGAACCGGCGGCGCAATGAGCAGTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Db 1141 GTGGAACCGGCGGCGCAATGAGCAGTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 Qy 1201 CCCCCACCCAGAGGCTGCTGAGAGTCAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 Db 1201 CCCCCACCCAGAGGCTGCTGAGAGTCAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 Qy 1261 TTGGCCGCAATCTGCAAGAGTGTGGCGGAGAGAGTGGCGCGAGAGTGTGGGA 1320  
 Db 1261 TTGGCCGCAATCTGCAAGAGTGTGGCGGAGAGAGTGGCGCGAGAGTGTGGGA 1320  
 Qy 1321 GGGCGGAGACACAGCTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 Db 1321 GGGCGGAGACACAGCTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 Qy 1381 CACCTCTGTGGGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
 Db 1381 CACCTCTGTGGGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
 Qy 1441 CCGGAGCGGAACCGGCTCTGTCTCCGATGAGAGTGTGGCTGGGCTGCGGCTGCGGCTG 1500  
 Db 1441 CCGGAGCGGAACCGGCTCTGTCTCCGATGAGAGTGTGGCTGGGCTGCGGCTGCGGCTG 1500

Db 1441 CCGGAGCGGAACCGGCTCTGTCCCAAGTGGAGTGTGGCTGGGCTGCGGCTGCGGCTG 1500  
 Qy 1501 TCTCCCAAGAGTCTGAGCTGGGGGTGAGAGTGTGGTCTACACAGGGGGCTATCTTCCC 1560  
 Db 1501 TCTCCCAAGAGTCTGAGCTGGGGGTGAGAGTGTGGTCTACACAGGGGGCTATCTTCCC 1560  
 Qy 1561 TTTTGGGAGCCCAACAGCGAG 1620  
 Db 1561 TTTTGGGAGCCCAACAGCGAG 1620  
 Qy 1621 CCCCTGCTCTCAGAGATATCAGTCAAGCTGTGTGCTCTCCAGTGTGCGGCGAGGCTG 1680  
 Db 1621 CCCCTGCTCTCAGAGATATCAGTCAAGCTGTGTGCTCTCCAGTGTGCGGCGAGGCTG 1680  
 Qy 1681 GTGGATGGCAAGATCTGTACCGTGAAGGAGTGGGGGCAACGAGTACTATGGCCACAG 1740  
 Db 1681 GTGGATGGCAAGATCTGTACCGTGAAGGAGTGGGGGCAACGAGTACTATGGCCACAG 1740  
 Qy 1741 GCGGAGGTAATCCAGAGAGCTGAGTCCCATATCAGCAATGATGTGCAATGGGCT 1800  
 Db 1741 GCGGAGGTAATCCAGAGAGCTGAGTCCCATATCAGCAATGATGTGCAATGGGCT 1800  
 Qy 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGC 1860  
 Db 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGC 1860  
 Qy 1861 ATTGATGCTGCAAGGCGGAG 1920  
 Db 1861 ATTGATGCTGCAAGGCGGAG 1920  
 Qy 1921 ACAGCAGTGGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980  
 Db 1921 ACAGCAGTGGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980  
 Qy 1981 AAGCAGGCGTCTACACAAAGTCAAGTCTTCCGAGAGTGAATCTTCCAGGCTATAAG 2040  
 Db 1981 AAGCAGGCGTCTACACAAAGTCAAGTCTTCCGAGAGTGAATCTTCCAGGCTATAAG 2040  
 Qy 2041 ACTCACTCCGAGCCAGCGGAGTGAAGCCAGTCTGAGCGGAGCTTCCGCTGCGGCA 2100  
 Db 2041 ACTCACTCCGAGCCAGCGGAGTGAAGCCAGTCTGAGCGGAGCTTCCGCTGCGGCA 2100  
 Qy 2101 GCGTCCAGGCGCGGAGTGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160  
 Db 2101 GCGTCCAGGCGCGGAGTGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160  
 Qy 2161 TTTTCTTCTTGGGCGGCTGCAAGAGTCAAGGACACCTTCCCTGCAAGGCTCTCTTTC 2220  
 Db 2161 TTTTCTTCTTGGGCGGCTGCAAGAGTCAAGGACACCTTCCCTGCAAGGCTCTCTTTC 2220  
 Qy 2221 CACAGTGGGCGGCGCACTGAGCCCGGAGCCAGCAACCTCACGCTCTGAGCCCAAGT 2280  
 Db 2221 CACAGTGGGCGGCGCACTGAGCCCGGAGCCAGCAACCTCACGCTCTGAGCCCAAGT 2280  
 Qy 2281 AAATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 2340  
 Db 2281 AAATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 2340  
 Qy 2341 AAATATTAAGATGTTTGAAT 2363  
 Db 2341 AAATATTAAGATGTTTGAAT 2363

RESULT 2  
 US-09-510-738A-188  
 / Sequence 188, Application US/09510738A  
 / Patent No. 6268165  
 / GENERAL INFORMATION:  
 / APPLICANT: O'Brien, Timothy J.  
 / TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of  
 / Ovarian Cancer  
 / FILE REFERENCE: D6223CIP-A  
 / CURRENT APPLICATION NUMBER: US/09/510, 738A



```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length cDNA of hepsin
US-09-861-966-188

Query Match      67.4%; Score 1593.6; DB 4; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY      736 CCCCCAGGCTCCCTCCCTCCATCTCTCAAGAGTCCACCTGGCCCCAGAGGTCA 795
      |||
DB      157 CCTGAGAGCTCCGCCCCCAGCTGCTGAGACCCCA-GGGTCCCACTGGCCCCAGAGGTCA 215
      |||

QY      796 GCCAGGGAGATTAACAAGAGGAGTACATGGGCGAGAGAGGGGGGCGGAGCTGTG 855
      |||
DB      216 GCCAGGGAGATTAACAAGAGGAGTACATGGGCGAGAGAGGGGGGCGGAGCTGTG 275
      |||

QY      856 CCATGCTGCTCAGACCCCAAGTGGAGCTCTCACTGGGAGAACCTGTCTACTTCTGACA 915
      |||
DB      276 CCATGCTGCTCAGACCCCAAGTGGAGCTCTCACTGGGAGAACCTGTCTACTTCTGACA 335
      |||

QY      916 GCCATCGGGGCGGATCTCTGGGCCATTGGCTGTTCTCTCAGAGTGAACAGAGCCG 975
      |||
DB      336 GCCATCGGGGCGGATCTCTGGGCCATTGGCTGTTCTCTCAGAGTGAACAGAGCCG 395
      |||

QY      976 CTGTACCCAGTGAAGGTGAGCTGCGGAGCGCTGGGCTCATGGTCTTTGAACAAGGAA 1035
      |||
DB      396 CTGTACCCAGTGAAGGTGAGCTGCGGAGCGCTGGGCTCATGGTCTTTGAACAAGGAA 455
      |||

QY      1036 GGGAGCTGAGCGGCTGTGTGCTCTCGGCTCAACCGCCAGAGGTAGCCGAGCTGACGTGC 1095
      |||
DB      456 GGGAGCTGAGCGGCTGTGTGCTCTCGGCTCAACCGCCAGAGGTAGCCGAGCTGACGTGC 515
      |||

QY      1096 GAGAGATGGGCTTCTCAAGGCACTGACCACTCCAGCTGAGAGCTGCAACCGCGGCGC 1155
      |||
DB      516 GAGAGATGGGCTTCTCAAGGCACTGACCACTCCAGCTGAGAGCTGCAACCGCGGCGC 575
      |||

QY      1156 GCCAATGGCACTGGGGCTTCTTCGTGTGAGACGAGGGGAGGCTGCCCAACCGAGAGG 1215
      |||
DB      576 GCCAATGGCACTGGGGCTTCTTCGTGTGAGACGAGGGGAGGCTGCCCAACCGAGAGG 635
      |||

QY      1216 CTGCTGAGAGTCACTCCGTGTGTGATTTGCCAGAGGCGCTTTCTTGCGCCGATCTGC 1275
      |||
DB      636 CTGCTGAGAGTCACTCCGTGTGTGATTTGCCAGAGGCGCTTTCTTGCGCCGATCTGC 695
      |||

QY      1276 CAAGACTGTGCGCGGAGAACTGCGCGTGAACCGCATCTGTGGAGGCGCGGAGACCAAC 1335
      |||
DB      696 CAAGACTGTGCGCGGAGAACTGCGCGTGAACCGCATCTGTGGAGGCGCGGAGACCAAC 755
      |||

QY      1336 TTGGAGCGGAGTGGCGGTGCAAGCTTGCCTATGATGAGGACACCTGTGTGGGGGA 1395
      |||
DB      756 TTGGAGCGGAGTGGCGGTGCAAGCTTGCCTATGATGAGGACACCTGTGTGGGGGA 815
      |||

QY      1396 TCCCTGTCTCTCCGGGGACTGGGTGTGACAGCCGCCCACTGCTTCCGAGACGGAACCGG 1455
      |||
DB      816 TCCCTGTCTCTCCGGGGACTGGGTGTGACAGCCGCCCACTGCTTCCGAGACGGAACCGG 875
      |||

QY      1456 GTCTGTCTCCGATGAGAGTGTGTCGCGGTGCGGTGCGGAGGCTCTCCCAAGGTCTG 1515
      |||
DB      876 GTCTGTCTCCGATGAGAGTGTGTCGCGGTGCGGTGCGGAGGCTCTCCCAAGGTCTG 935
      |||

QY      1516 CAGCTGGGGGTGAGAGCTGTGTCTAACACGAGGGGCTATCTTCCCTTTGGGAGACCCCAAC 1575
      |||
DB      936 CAGCTGGGGGTGAGAGCTGTGTCTAACACGAGGGGCTATCTTCCCTTTGGGAGACCCCAAC 995
      |||

QY      1576 AGCAGAGGAACAAGCAAGATATTGCCCTGTATCAACCTTCCAGTCCCTGCCCCCTCA 1635
      |||
DB      996 AGCAGAGGAACAAGCAAGATATTGCCCTGTATCAACCTTCCAGTCCCTGCCCCCTCA 1055
      |||

QY      1636 GAATATCATCCAGCTGTGTGCTCCAGCTGCGGCGAGGCGCTGTGTGATGAGCAAGATC 1695
      |||
DB      1056 GAATATCATCCAGCTGTGTGCTCCAGCTGCGGCGAGGCGCTGTGTGATGAGCAAGATC 1115
      |||
```

```

QY      1696 TGTAACGTGACGGGCTGGGCAACAAGAGTATATGATGAGCCAAAGGCGGGGTATCTCAG 1755
      |||
DB      1116 TGTAACGTGACGGGCTGGGCAACAAGAGTATATGATGAGCCAAAGGCGGGGTATCTCAG 1175
      |||

QY      1756 GAGGCTGAGTCCCATTAATCAAGATGATGTCTGCAATGAGCGCTGACTTCTATGAAAC 1815
      |||
DB      1176 GAGGCTGAGTCCCATTAATCAAGATGATGTCTGCAATGAGCGCTGACTTCTATGAAAC 1235
      |||

QY      1816 CAATTAAGCCCAAGATGTTCTGTGCTGGGTACCCGAGGGGTGGCAATTGATGCTGACAG 1875
      |||
DB      1236 CAATTAAGCCCAAGATGTTCTGTGCTGGGTACCCGAGGGGTGGCAATTGATGCTGACAG 1295
      |||

QY      1876 GCGACAGCGGTGGTCCCTTTGTGTGTGAGAGACAGACTCTCTGAGAGCCACGTGTGCGG 1935
      |||
DB      1296 GCGACAGCGGTGGTCCCTTTGTGTGTGAGAGACAGACTCTCTGAGAGCCACGTGTGCGG 1355
      |||

QY      1936 CTGTGTGACATTGTGATTTGGGAGCATGAGCTGTGCTGCTGGCCCCAGAACCGAGCTTAC 1995
      |||
DB      1356 CTGTGTGACATTGTGATTTGGGAGCATGAGCTGTGCTGCTGGCCCCAGAACCGAGCTTAC 1415
      |||

QY      1996 ACCAATGCTAGTACTTCCGGGAGTGGATCTTCAGGCGCATTAAGACTCATCTCCAGAGCC 2055
      |||
DB      1416 ACCAATGCTAGTACTTCCGGGAGTGGATCTTCAGGCGCATTAAGACTCATCTCCAGAGCC 1475
      |||

QY      2056 AGCGCATGTGTGACCCAGCTCTGACCGGTGCTTCTGCTGCGCAGGCTCCAGGGCCGA 2115
      |||
DB      1476 AGCGCATGTGTGACCCAGCTCTGACCGGTGCTTCTGCTGCGCAGGCTCCAGGGCCGA 1535
      |||

QY      2116 GGTGATCCCGGTGGTGGATTCACGCTGGGCCAGATGGAGGTTTCTTCTTGAGGCC 2175
      |||
DB      1536 GGTGATCCCGGTGGTGGATTCACGCTGGGCCAGATGGAGGTTTCTTCTTGAGGCC 1595
      |||

QY      2176 CGGTCCACAGTGTCAAGGACACCTCCCTCAGGGTCTCTTCCACAGTGGCGGGCC 2235
      |||
DB      1596 CGGTCCACAGTGTCAAGGACACCTCCCTCAGGGTCTCTTCCACAGTGGCGGGCC 1655
      |||

QY      2236 ACTAGCCCCGAGACCAACCACTGACCTCTCTGACCCCAATGTAATTTGTTCTGCT 2295
      |||
DB      1656 ACTAGCCCCGAGACCAACCACTGACCTCTCTGACCCCAATGTAATTTGTTCTGCT 1715
      |||

QY      2296 GTCTGGAGCTCTGTCTTAAAGTGTGCTGATGATGAGTGTCTTTTAAATTAATAAGTGG 2355
      |||
DB      1716 GTCTGGAGCTCTGTCTTAAAGTGTGCTGATGATGAGTGTCTTTTAAATTAATAAGTGG 1775
      |||

QY      2356 TTTTGATT 2363
      |||
DB      1776 TTTTGATT 1783
      |||

RESULT 4
US-09-919-048-188
; Sequence 188, Application US/09919048
; Patent No. 6787354
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Cannon, Martin J.
; APPLICANT: Santin, Alessandro
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
; FILE REFERENCE: D6223CIP/A/D/CIP
; CURRENT APPLICATION NUMBER: US/09/919,048
; PRIOR APPLICATION NUMBER: 2001-07-30
; PRIOR FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 188
; SEQ ID NO 188
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length cDNA of hepsin
US-09-919-048-188

Query Match      67.4%; Score 1593.6; DB 4; Length 1783;
```



Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCCAGGCGCTGCTCCCGCTGCATCTCTCAGAGTCCACCTGCGCCAGAGGTCA 795  
Db 157 CTTCCAGGCTCCGCCCCCACTGTGAGCCCA-GGGTCCCACTCCGCGCCAGAGGTCA 215  
QY 796 GCCAGGGAATCATTAACAAGAGGAGTGAACATGGGGGCAAGAGAGAGGTGGCCGAGCTGTG 855  
Db 216 GCCAGGGAATCATTAACAAGAGGAGTGAACATGGGGGCAAGAGAGAGGTGGCCGAGCTGTG 275  
QY 856 CCATGCTGCTCCAGAGCCCAAGGTGAGCACTCTCACTGGGGGAGCCCTGCTACTTTGACA 915  
Db 276 CCAATGCTGCTCCAGAGCCCAAGGTGAGCACTCTCACTGGGGGAGCCCTGCTACTTTGACA 335  
QY 916 GCCATCGGGGCGGCATCTCTGGGCGCATTTGTGCTGTTCTCTCAGAGTGAACAGAGCGG 975  
Db 336 GCCATCGGGGCGGCATCTCTGGGCGCATTTGTGCTGTTCTCTCAGAGTGAACAGAGCGG 395  
QY 976 CTGTATCCAGTGAAGGTGAGCTGTGAGCGGAGCGCTCGGCTCATGTGTTTGAACAAGCGAA 1035  
Db 396 CTGTATCCAGTGAAGGTGAGCTGTGAGCGGAGCGCTCGGCTCATGTGTTTGAACAAGCGAA 455  
QY 1036 GGGAGGTGGCGGCTGCTGCTGCTCCGCTCAGAGCGGAGGAGGAGGAGGAGGAGGAGG 1095  
Db 456 GGGAGGTGGCGGCTGCTGCTGCTCCGCTCAGAGCGGAGGAGGAGGAGGAGGAGGAGG 515  
QY 1096 GAGAGAGTGGGCTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1155  
Db 516 GAGAGAGTGGGCTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 575  
QY 1156 GCCATAGGCAAGTGGGCTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1215  
Db 576 GCCATAGGCAAGTGGGCTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 635  
QY 1216 CTGCTGAGGTGATCTCCGCTGTGATTTGCCAGAGGCGCTTCTTGGCGGATCTGCG 1275  
Db 636 CTGCTGAGGTGATCTCCGCTGTGATTTGCCAGAGGCGCTTCTTGGCGGATCTGCG 695  
QY 1276 CAAAGACTGTGGCGGAGAGAGCTGCGGTGAGCGGATGTTGGAGGCGGAGAGAGCAGC 1335  
Db 696 CAAAGACTGTGGCGGAGAGAGCTGCGGTGAGCGGATGTTGGAGGCGGAGAGAGCAGC 755  
QY 1336 TTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1395  
Db 756 TTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 815  
QY 1396 TCCCTGCTCTCCGAGGAGCTGGGTGTGAGCGGCGGCTGCTTCCGAGAGGAGAGCGG 1455  
Db 816 TCCCTGCTCTCCGAGGAGCTGGGTGTGAGCGGCGGCTGCTTCCGAGAGGAGAGCGG 875  
QY 1456 GTCTGTGCTCCGAGTGGGTGTTTGGCGGCTGCGGTGAGCGGCGGCTTCCGAGAGT 1515  
Db 876 GTCTGTGCTCCGAGTGGGTGTTTGGCGGCTGCGGTGAGCGGCGGCTTCCGAGAGT 935  
QY 1516 CAGCTGGGGGTGAGGCTGTGTGCTTCAACGCGGGGCTTCTTCCCTTTCCGAGAGCC 1575  
Db 936 CAGCTGGGGGTGAGGCTGTGTGCTTCAACGCGGGGCTTCTTCCCTTTCCGAGAGCC 995  
QY 1576 AGCAGAGAGAGAGAGAGATATTGGCTTCCAGCTTCCAGGCTTCCGAGAGCCCTCACA 1635  
Db 996 AGCAGAGAGAGAGAGAGATATTGGCTTCCAGCTTCCAGGCTTCCGAGAGCCCTCACA 1055  
QY 1636 GAATATCATTCAGCTGTGTGCTTCCAGCTTCCAGGCTTCCGAGAGCCCTCACA 1695  
Db 1056 GAATATCATTCAGCTGTGTGCTTCCAGCTTCCAGGCTTCCGAGAGCCCTCACA 1115  
QY 1696 TGTATCCGTGAGAGGCTGGGGGAGAGAGAGTATATGAGGAGAGGCGGGGTACTCCAG 1755  
Db 1116 TGTATCCGTGAGAGGCTGGGGGAGAGAGAGTATATGAGGAGAGGCGGGGTACTCCAG 1175  
QY 1756 GAGAGCTGAGTCCCATTAATCAGAGATGATGTGAGAGTGGCGGCTTCAATGAAGC 1815

Db 1176 GAGGCTGAGTCCCATTAATCAGAGATGATGTGAGAGTGGCGGCTGATCTTATGAGAAC 1235  
QY 1816 CAGATCAAGGCCCAAGATGTTTGTGTGTGCTTACCCCGAGGAGGATGATGATGCTGCGAG 1875  
Db 1236 CAGATCAAGGCCCAAGATGTTTGTGTGTGCTTACCCCGAGGAGGATGATGATGCTGCGAG 1295  
QY 1876 GGGCGAGCGGTTGGCTTGTGTGTGAGGAGAGAGATCTCTCGAGAGCGGAGTGGGCGG 1935  
Db 1296 GGGCGAGCGGTTGGCTTGTGTGTGAGGAGAGAGATCTCTCGAGAGCGGAGTGGGCGG 1355  
QY 1936 CTGTGTGAGTGTGAGTGTGGGAGAGTGTGTGTGCTTGTGTGTGCTTGTGTGTGCTTGT 1995  
Db 1356 CTGTGTGAGTGTGAGTGTGGGAGAGTGTGTGTGCTTGTGTGTGCTTGTGTGTGCTTGT 1415  
QY 1996 ACCAAGTCAATGATCTTCCGAGAGTGTGTGTGTGCTTGTGTGTGCTTGTGTGTGCTT 2055  
Db 1416 ACCAAGTCAATGATCTTCCGAGAGTGTGTGTGTGCTTGTGTGTGCTTGTGTGTGCTT 1475  
QY 2056 AGCGGATGTGAGAGCCAGCTGTGAGCGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 2115  
Db 1476 AGCGGATGTGAGAGCCAGCTGTGAGCGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 1535  
QY 2116 GGTATCCCGGT 2175  
Db 1536 GGTATCCCGGT 1595  
QY 2176 CGGTTCACAGGTCCAGAGAGACCTCTCTCCAGGAGTCTCTTTCACAGTGGCGGCGC 2235  
Db 1596 CGGTTCACAGGTCCAGAGAGACCTCTCTCCAGGAGTCTCTTTCACAGTGGCGGCGC 1655  
QY 2236 ACTAGGCGCGGAG 2295  
Db 1656 ACTAGGCGCGGAG 1715  
QY 2296 GTCTGAGAGTCTCTTCTAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2355  
Db 1716 GTCTGAGAGTCTCTTCTAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1775  
QY 2356 TTTTGATT 2363  
Db 1776 TTTTGATT 1783

RESULT 5  
US-09-949-016-1311  
; Sequence 1311, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1311  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-1311

Query Match 67.4%; Score 1593.6; DB 4; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCCAGGCGCTGCTCCCGTGCATCTCTCAGAGTCCACCTGCGCCAGAGGTCA 795

|    |      |   |      |
|----|------|---|------|
| Db | 157  | CTTGAGAGCTTCGCCCCCACTGTGACCCCA -GGGTCCACCTTGCCAGGAGGTCA         | 215  |
| Qy | 796  | GCCAGGGAAATCATTTAACAGAGGACAGTGCATGCGCGAGAAAGAGGTGTGCGGACTGTG    | 855  |
| Db | 216  | GCCAGGGAAATCATTTAACAGAGGACAGTGCATGCGCGAGAAAGAGGTGTGCGGACTGTG    | 275  |
| Qy | 856  | CCATGCTGTCTCAGACCCCAAGTGTGACGCTCTCACTGTGCGGAGACCTGTCTACTTTGACA  | 915  |
| Db | 276  | CCATGCTGTCTCAGACCCCAAGTGTGACGCTCTCACTGTGCGGAGACCTGTCTACTTTGACA  | 335  |
| Qy | 916  | GCCATGCGGGGGGGGATCCGTGGGCCAATTGTGTGTCTCTCCAGAGTGAACAAGAGCG      | 975  |
| Db | 336  | GCCATGCGGGGGGGGATCCGTGGGCCAATTGTGTGTCTCTCTCAGAGTGAACAAGAGCG     | 395  |
| Qy | 976  | CTGTACCCAGTGCAGGTACGCTCTGCGGACGCTCGGCTCATGTCTTTTGAACAAGCGAA     | 1035 |
| Db | 396  | CTGTACCCAGTGCAGGTACGCTCTGCGGACGCTCGGCTCATGTCTTTTGAACAAGCGAA     | 455  |
| Qy | 1036 | GGGACGTGTGGCGGCTGTGTGTCTCTCGCGCTCCAA CGCAGAGGTAGCCGACTGCTGC     | 1095 |
| Db | 456  | GGGACGTGTGGCGGCTGTGTGTCTCTCGCGCTCCAA CGCAGAGGTAGCCGACTGCTGC     | 515  |
| Qy | 1096 | GAGAGATGTGGCTTCTTCAGGGGACGTACCCACTTCCAGCTGTGAGAGTGTGCAACGGCGGGC | 1155 |
| Db | 516  | GAGAGATGTGGCTTCTTCAGGGGACGTACCCACTTCCAGCTGTGAGAGTGTGCAACGGCGGGC | 575  |
| Qy | 1156 | GCCATATGCAAGTGGGCTTCTGTGTGTGACGAGGGGAGGAGCTGCCCCA CACCAGAG      | 1215 |
| Db | 576  | GCCATATGCAAGTGGGCTTCTGTGTGTGACGAGGGGAGGAGCTGCCCCA CACCAGAG      | 635  |
| Qy | 1216 | CTGTGTGAGGTCACTCTCCGTGTGTGATTTGCCCAAGGCGGTTTCTTGAGCCGACTGTGC    | 1275 |
| Db | 636  | CTGTGTGAGGTCACTCTCCGTGTGTGATTTGCCCAAGGCGGTTTCTTGAGCCGACTGTGC    | 695  |
| Qy | 1276 | CAAGACTGTGGCCCCGAGAACGTGCGGTGGAACCGCATCTGTGTGAGAGCGCGGGACACCAAC | 1335 |
| Db | 696  | CAAGACTGTGGCCCCGAGAACGTGCGGTGGAACCGCATCTGTGTGAGAGCGCGGGACACCAAC | 755  |
| Qy | 1336 | TTTGGGCGGCTGGCGCGGTGCAAGTCAAGCTTGCTATATGATGAGCAACCTCTGTGTGGGGA  | 1395 |
| Db | 756  | TTTGGGCGGCTGGCGCGGTGCAAGTCAAGCTTGCTATATGATGAGCAACCTCTGTGTGGGGA  | 815  |
| Qy | 1396 | TCCCTGTCTCTCGGAGACTGTGGGTGTGACAGCCGCGCCACTCTTCCCGAGCGGAACCGG    | 1455 |
| Db | 816  | TCCCTGTCTCTCGGAGACTGTGGGTGTGACAGCCGCGCCACTCTTCCCGAGCGGAACCGG    | 875  |
| Qy | 1456 | GTCTGTCTCCGATGCGAGTGTGTTGCCGTGTGCGGTGCGCGAGGCTCTTCCCA CGGTCTG   | 1515 |
| Db | 876  | GTCTGTCTCCGATGCGAGTGTGTTGCCGTGTGCGGTGCGCGAGGCTCTTCCCA CGGTCTG   | 935  |
| Qy | 1516 | CAGCTGGGGGGTGCAGGCTGTGTGTCTACACAGGGGGCTATCTTCCCTTTGGGACCCCAAC   | 1575 |
| Db | 936  | CAGCTGGGGGGTGCAGGCTGTGTGTCTACACAGGGGGCTATCTTCCCTTTGGGACCCCAAC   | 995  |
| Qy | 1576 | AGCAGAGGAACAGCAACGATATTGCTCCGTGTCACTCTCAAGTCCCTGCCCCCTCA        | 1635 |
| Db | 996  | AGCAGAGGAACAGCAACGATATTGCTCCGTGTCACTCTCAAGTCCCTGCCCCCTCA        | 1055 |
| Qy | 1636 | GAATPACATCCAGCCTGTGTGCTTCCAGCTGCGGCGCAGGCGCTGTGTGTGATGTGCAAGATC | 1695 |
| Db | 1056 | GAATPACATCCAGCCTGTGTGCTTCCAGCTGCGGCGCAGGCGCTGTGTGTGATGTGCAAGATC | 1115 |
| Qy | 1696 | TGTACCCGTGAACGGGCTGTGGGCAACCGAGATATCTATGGCCAAACAGGCGGGGTACTCCAG | 1755 |
| Db | 1116 | TGTACCCGTGAACGGGCTGTGGGCAACCGAGATATCTATGGCCAAACAGGCGGGGTACTCCAG | 1175 |
| Qy | 1756 | GAGGCTCGAGTCCCATATACGCAATATGATGTCTGCAATGTGCGGCTGACTTCTATGTGAAC  | 1815 |
| Db | 1176 | GAGGCTCGAGTCCCATATACGCAATATGATGTCTGCAATGTGCGGCTGACTTCTATGTGAAC  | 1235 |
| Qy | 1816 | CAGATCAAGCCCAAGATGTCTGTGTGTGCTAACCCCGAGGGGTGCATTGATGTCTGCCAG    | 1875 |

|    |      |  |       |
|----|------|--|-------|
| Db | 1236 | CAGATCAAGCCCAAGATGTTCTGTGCTGGGATACCCGAGGGGTGCATTGATGCTGCGAAG     | 12935 |
| Qy | 1876 | GGCGCAGACGGGTGGTCCCTTTGTGTGTGAGAGACAGCATCTCTCGAGGCCACGTTGGCGG    | 19335 |
| Db | 1296 | GGCGCAGACGGGTGGTCCCTTTGTGTGTGAGAGACAGCATCTCTCGAGGCCACGTTGGCGG    | 13555 |
| Qy | 1936 | CTGTGTGGCATTGTGATTTGGGGGACAGTGGCTGTGCCCTGGGCCGAGAACGACAGCGTCTAC  | 19955 |
| Db | 1356 | CTGTGTGGCATTGTGAGTTGGGGGACAGTGGCTGTGCCCTGGGCCGAGAACGACAGCGTCTAC  | 14155 |
| Qy | 1996 | ACCAAAGTCAGTGACTTCCGGGAGTGTGATCTTCAGAGCCATTAAGACTCACTCCGAAGCC    | 20555 |
| Db | 1416 | ACCAAAGTCAGTGACTTCCGGGAGTGTGATCTTCAGAGCCATTAAGACTCACTCCGAAGCC    | 14755 |
| Qy | 2056 | AGCGGCATTGTGACCCAGCCAGCTCTGA                                     | 21155 |
| Db | 1476 | AGCGGCATTGTGACCCAGCCAGCTCTGA                                     | 15355 |
| Qy | 2116 | GGTATATCCCGGTGGTGGGATTCACAGCTGTGGGCCAGAGATGGGACGTTTTCTTCTTGGGGCC | 21755 |
| Db | 1536 | GGTATATCCCGGTGGTGGGATTCACAGCTGTGGGCCAGAGATGGGACGTTTTCTTCTTGGGGCC | 15955 |
| Qy | 2176 | CGGTCCACAGGTCCAAAGACACCCCTCCCTCAGAGGTCTCTCTTCCACAGTGGCGGGCCC     | 22355 |
| Db | 1596 | CGGTCCACAGGTCCAAAGACACCCCTCCCTCAGAGGTCTCTCTTCCACAGTGGCGGGCCC     | 16555 |
| Qy | 2236 | ACTCAGACCCCGAGACCAACCACTCAACCTCTCGACCCCCATGTAAATATTTGTTCTGCT     | 22955 |
| Db | 1656 | ACTCAGACCCCGAGACCAACCACTCAACCTCTCGACCCCCATGTAAATATTTGTTCTGCT     | 17155 |
| Qy | 2296 | GTCTGGGACTCTCTGTAGGTGCCCCCTGATGATGGGATGCTCTTTAAATATAAAGATGG      | 23555 |
| Db | 1716 | GTCTGGGACTCTCTGTAGGTGCCCCCTGATGATGGGATGCTCTTTAAATATAAAGATGG      | 17755 |
| Qy | 2356 | TTTTTGATT 2363   |       |
| Db | 1776 | TTTTTGATT 1783   |       |

```

RESULT 6
US-09-820-002-1
; Sequence 1, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1:001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1615
; TYPE: DNA
; ORGANISM: HUMAN
US-09-820-002-1

```

| Query Match                | 53.7% | Score 1269;   | DB 4;   | Length 1615; |
|----------------------------|-------|---|---------|--------------|
| Best Local Similarity      | 91.1% | Pred. No. 2.7e-308;                                       |         |              |
| Matches 1411, Conservative | 0;    | Indels 15;  | Gaps 1; |              |
| QY                         | 815   | GAGCAGCGAATATGCCGAGAAGAGGTGGCCGAGCTGTGCCATGTGCTCCAGACCA   | 874     |              |
|                            |       |   |         |              |
| Db                         | 167   | GTGGCCCGACATGATGTGTTTTCAGATGCGCGACTGTGCCATGCTGCCAGACCA    | 226     |              |
| QY                         | 875   | AGGTGGACGCTCTACATCCGGGGACCCGCGACTTTTGACAGCCATCGGGCGGATCCT | 934     |              |
| Db                         | 227   | AGGTGGACGCTCTACATCCGGGGACCCGCTACTTTTGACAGCCATCGGGCGGATCCT | 286     |              |

935 GGGCCATTGTGCTTTCTCTCAGAGAGACAGAGCCGCTGTAACCAAGTCAGGTC 994  
Db GGGCCATTGTGCTTTCTCTCAGAGAGACAGAGCCGCTGTAACCAAGTCAGGTC 346  
995 GCTCTGCGGAGCGCTGCTCATGTCTTTGACAAAGAGAGAGAGCTGCGCTGCTGT 1054  
Db GCTCTGCGGAGCGCTGCTCATGTCTTTGACAAAGAGAGAGAGCTGCGCTGCTGT 406  
1055 GCTCTGCGGAGCGCTGCTCATGTCTTTGACAAAGAGAGAGAGCTGCGCTGCTGT 1114  
Db GCTCTGCGGAGCGCTGCTCATGTCTTTGACAAAGAGAGAGAGCTGCGCTGCTGT 466  
1115 GGGCACTAGCCCACTCCAGAGCTGAGAGCTGCGAGAGAGAGAGAGAGAGAGCT 1174  
Db GGGCACTAGCCCACTCCAGAGCTGAGAGCTGCGAGAGAGAGAGAGAGAGAGCT 466  
1175 TCTTGTGTGTGAG 1234  
Db TCTTGTGTGTGAG 466  
1235 TGTGTGATTTGAG 1294  
Db TGTGTGATTTGAG 467  
1295 AGCTGCGGAG 1354  
Db AGCTGCGGAG 583  
1355 AAGTCAGGCTTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1414  
Db AAGTCAGGCTTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643  
1415 GAGTGTGAG 1474  
Db GAGTGTGAG 703  
1475 TGTGTGCGGAG 1534  
Db TGTGTGCGGAG 763  
1535 TGTGTGAG 1594  
Db TGTGTGAG 823  
1595 ATATTGCGGAG 1654  
Db ATATTGCGGAG 883  
1655 GCTTGTGAG 1714  
Db GCTTGTGAG 943  
1715 GCAACAG 1774  
Db GCAACAG 1003  
1775 TCAGCAATGATGTGTGCAATGTGCGCTGATCTTATGAAACAGATCAAGCCAGAT 1834  
Db TCAGCAATGATGTGTGCAATGTGCGCTGATCTTATGAAACAGATCAAGCCAGAT 1063  
1835 TCTGTGTGCTGAG 1894  
Db TCTGTGTGCTGAG 1123  
1895 TGTGTGTGAG 1954  
Db TGTGTGTGAG 1183  
1955 GGGGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2014  
Db GGGGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243

2015 GGGAGTGATCTTCAGGCGCATTAAGACTCACTCCGAAAGCCAGGCGGATGAGACCG 2074  
Db GGGAGTGATCTTCAGGCGCATTAAGACTCACTCCGAAAGCCAGGCGGATGAGACCG 1303  
2075 TCTGACCGGAGCTTCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2134  
Db TCTGACCGGAGCTTCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1363  
2135 TCCAGCTGAG 2194  
Db TCCAGCTGAG 1423  
2195 CACCTGCTGAG 2254  
Db CACCTGCTGAG 1483  
2255 CAACCTGAG 2314  
Db CAACCTGAG 1543  
2315 GTGCGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2363  
Db GTGCGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1592

## RESULT 7

US-09-000-846-1  
Sequence 1, Application US/09000846  
Patent No. 5981830  
GENERAL INFORMATION:  
APPLICANT: WU, QINGYU  
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH  
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MILLER, WHITE, ZELANO & BRANIGAN, P. C.  
STREET: 2200 CLARENDON BLVD, SUITE 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: US  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/000,846  
FILING DATE: 30-DEC-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/866,058  
FILING DATE: 30-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: LEBOVITZ, RICHARD M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: BERLX 65P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-243-6333  
TELEFAX: 703-243-6410  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1605 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
FEATURE:



```

Query Match      34.1%; Score 806.8; DB 4; Length 30337;
Best Local Similarity 99.8%; Pred. No. 7.6e-192;
Matches 808; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 35 CCACAGGTGAGGACAGCTGCTAGCAGGCCCCCAGGCCCCCAGGCTCTGCTCCAGGCGC 94
DB 3211 CCTCAGGTGAGGACAGCTGCTAGCAGGCCCCCAGGCCCCCAGGCTCTGCTCCAGGCGC 3270
QY 95 CCGCTGCTGCGGGGCGACCATCTCTCTCCAGGCGCTGAGACTGACCCGCGGAC 154
DB 3271 CCGCTGCTGCGGGGCGACCATCTCTCTCCAGGCGCTGAGACTGACCCGCGGAC 3330
QY 155 TACCTGAGGCTCCGCGCGCCAGCTGAGACCCCAAGGTAAGAGAGAGGCGCCCAAGT 214
DB 3331 TACCTGAGGCTCCGCGCGCCAGCTGAGACCCCAAGGTAAGAGAGAGGCGCCCAAGT 3390
QY 215 CACAGTTCCAGCCCTGAGAGACAGGGGTTCCCTCATCCCCCAGCCAGCTTAATGCGCAC 274
DB 3391 CACAGTTCCAGCCCTGAGAGACAGGGGTTCCCTCATCCCCCAGCCAGCTTAATGCGCAC 3450
QY 275 TCTTAATAGAGGGGTTCTGAGGACCTGAAGAGGGGCACTATGACGTCTCCCAAGCAC 334
DB 3451 TCTTAATAGAGGGGTTCTGAGGACCTGAAGAGGGGCACTATGACGTCTCCCAAGCAC 3510
QY 335 CTAGGTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 394
DB 3511 CTAGGTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3570
QY 395 CCAGAGCCAGAGAGTTCCAGGCTCAGGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 454
DB 3571 CCAGAGCCAGAGAGTTCCAGGCTCAGGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCT 3630
QY 455 CCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 514
DB 3631 CCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3690
QY 515 CCGGATCTCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 574
DB 3691 CCGGATCTCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3750
QY 575 TCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 634
DB 3751 TCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3810
QY 635 GCGTCGATGATCAGCGGCTGCTCTGAGCAAGGCCAGTCCCTACAGCTGCTGAGTGA 694
DB 3811 GCGTCGATGATCAGCGGCTGCTCTGAGCAAGGCCAGTCCCTACAGCTGCTGAGTGA 3870
QY 655 CCGCTGAGACTGAGGCGCGCGCAGAGACTGAGGCTGAGGCTCTCTCTCTCTCTCTCTCT 754
DB 3871 CCGCTGAGACTGAGGCGCGCGCAGAGACTGAGGCTGAGGCTCTCTCTCTCTCTCTCTCT 3930
QY 755 CGTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 814
DB 3931 CGTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3990
QY 815 GAGGAGTGCATGCGCGCAGAGAGAGAGGAGT 844
DB 3991 GAGGAGTGCATGCGCGCAGAGAGAGAGGAGT 4020

RESULT 9
US-09-280-116-71/c
; Sequence 71, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 71
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-71

Query Match      16.6%; Score 391.2; DB 3; Length 614;
Best Local Similarity 88.9%; Pred. No. 2.7e-88;
Matches 537; Conservative 0; Mismatches 53; Indels 14; Gaps 10;

QY 1774 ATCAGCATGATGCTGCTGAATGCGCTGAGCTTATGGA-AACAGATCAAG-CCCAAGA 1831
DB 614 ATCAGCATGATGCTGCTGAATGCGCTGAGCTTATGGA-AACAGATCAAG-CCCAAGA 555
QY 1832 TGTTCGTGCTGAGCTACCCGAGGCTGAGTATG-ATGCTGCGAGGCGAGAGCGGTGT 1890
DB 554 TGTTCGTGCTGAGCTACCCGAGGCTGAGTATG-ATGCTGCGAGGCGAGAGCGGTGT 495
QY 1891 CCGTTTGTGTGAGAGACAGCATCTCTGAGACGCCACGCTTGCGGCTGTGTGCAATGTG 1950
DB 494 CCGTTTGTGTGAGAGACAGCATCTCTGAGACGCCACGCTTGCGGCTGTGTGCAATGTG 435
QY 1951 AGTTGG-AGCATGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2009
DB 434 AGTTGGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375
QY 2010 CTTCCGAGGAGTGAATCTTCCAGGCGCATTAAGACT--CACTCGAAGCCAGGCGG-ATGCT 2066
DB 374 CTTCCGAGGAGTGAATCTTCCAGGCGCATTAAGACT--CACTCGAAGCCAGGCGG-AT 315
QY 2067 GACCAAGCTGACCGG--TGCTTCTGCTGCGAGCTCTGAGGCGCGAGGTGATCC 2123
DB 314 GACCAAGCTGACCGG--TGCTTCTGCTGCGAGCTCTGAGGCGCGAGGTGATCC 255
QY 2124 CCGTGTGTGATTCACAGCT-AGGCGAGAGATGAGACGTTTCTTCTGCGGCGCGGTGCA 2182
DB 254 CCGTGTGTGATTCACAGCT-AGGCGAGAGATGAGACGTTTCTTCTGCGGCGCGGTGCA 195
QY 2183 CAGGTCCAGAGACACCCCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2242
DB 194 CAGGTCCAGAGACACCCCTCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 135
QY 2243 CCGAGACCAACCAACCTCCTCTGAGCCCGCATGTAATATGTTCTGCT-GTCTGG 2301
DB 134 CCGAGACCAACCAACCTCCTCTGAGCCCGCATGTAATATGTTCTGCT-GTCTGAG 75
QY 2302 GACTTCCTGTCTAGGTGCCCC--TGATGATGAGATGCTCTTTAAATATTAAGATGTTT 2359
DB 74 GACTTCCTGTCTAGGTGCCCC--TGATGATGAGATGCTCTTTAAATATTAAGATGTTT 15
QY 2360 GATT 2363
DB 14 GATT 11

RESULT 10
US-09-820-002-3
; Sequence 3, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: C1001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0

```

SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 21784  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)---(21784)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-820-002-3

Query Match 13.8%; Score 325; DB 4; Length 21784;  
Best Local Similarity 100.0%; Pred. No. 4,7e-71;  
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2039 AGACTCATCCGAGACCGCCGATGATGACCCAGCTCTGACCGGTCTTCTGCTCG 2098  
DB 19460 AGACTCATCCGAGACCGCCGATGATGACCCAGCTCTGACCGGTCTTCTGCTCG 19519  
QY 2099 CAGCCTCAGGGCCCGAGGTGATCCGGTGGTGGATCAGCGTGGCGAGATGGAGC 2158  
DB 19520 CAGCCTCAGGGCCCGAGGTGATCCGGTGGTGGATCAGCGTGGCGAGATGGAGC 19579  
QY 2159 GTTTTCTCTTCTGCGCCCGATCCAGGTCCAGGACACCTTCCCTCCAGGGTCTCT 2218  
DB 19580 GTTTTCTCTTCTGCGCCCGATCCAGGTCCAGGACACCTTCCCTCCAGGGTCTCT 19639  
QY 2219 TCACAGTGGGGGCGCCACTCAGCCCGAGACACCCCAACTCAGCTCTGACCCCAT 2278  
DB 19640 TCACAGTGGGGGCGCCACTCAGCCCGAGACACCCCAACTCAGCTCTGACCCCAT 19699  
QY 2279 GTAATATATGTTCTGCTGTGAGGATCTCTGTCTAGTGGCCCTGATGATGGATGCT 2338  
DB 19700 GTAATATATGTTCTGCTGTGAGGATCTCTGTCTAGTGGCCCTGATGATGGATGCT 19759  
QY 2339 TTAATAATAAAGATGTTTGAAT 2363  
DB 19760 TTAATAATAAAGATGTTTGAAT 19784

RESULT 11  
US-09-820-002-15  
; Sequence 15, Application US/09820002  
; Patent No. 6482630  
; GENERAL INFORMATION:  
; APPLICANT: Gam, Weiniu  
; APPLICANT: Ye, Jane  
; APPLICANT: DiFrancesco, Valentina  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: C1001194  
; CURRENT APPLICATION NUMBER: US/09/820.002  
; CURRENT FILING DATE: 2001-03-29  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-820-002-15

Query Match 8.1%; Score 192; DB 4; Length 601;  
Best Local Similarity 90.3%; Pred. No. 2.5e-38;  
Matches 204; Conservative 1; Mismatches 21; Indels 0; Gaps 0;  
QY 1411 GACTGGGTCTGACGACGCGCCACTGCTTCCGAGCGGAAACGGGTCTGTCCGATGG 1470  
DB 173 GCGAGGTGGACACCTCCACCCCTTCCCTGATGAGCGGAAACGGGTCTGTCCGATGG 232  
QY 1471 CGAGTGTTCGCGGTGCGGTGGCCGAGGCTCTCTCCCAAGCTGTGCACTGGGGGTGAG 1530

DB 233 CGAGTGTTCGCGGTGCGGTGGCCGAGGCTCTCTCCCAAGCTGTGCACTGGGGGTGAG 292  
QY 1531 GCTGTGTCTACACACGCGGGGCTATCTTCCCTTGGGAGCCCAACGAGGAGAAACAGC 1590  
DB 293 GCTGTGTCTACACACGCGGGGCTATCTTCCCTTGGGAGCCCAACGAGGAGAAACAGC 352  
QY 1591 AACGATATTGCGCTGTGTCACCTCTCCAGTCCCTGCCCCCTACAG 1636  
DB 353 AACGATATTGCGCTGTGTCACCTCTCCAGTCCCTGCCCCCTACAG 398

RESULT 12  
US-09-949-016-45925  
; Sequence 45925, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949.016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45925  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-45925

Query Match 8.1%; Score 192; DB 4; Length 601;  
Best Local Similarity 90.3%; Pred. No. 2.5e-38;  
Matches 204; Conservative 1; Mismatches 21; Indels 0; Gaps 0;

QY 1411 GACTGGGTCTGACGACGCGCCACTGCTTCCGAGCGGAAACGGGTCTGTCCGATGG 1470  
DB 173 GCGAGGTGGACACCTCCACCCCTTCCCTGATGAGCGGAAACGGGTCTGTCCGATGG 232  
QY 1471 CGAGTGTTCGCGGTGCGGTGGCCGAGGCTCTCTCCCAAGCTGTGCACTGGGGGTGAG 1530  
DB 233 CGAGTGTTCGCGGTGCGGTGGCCGAGGCTCTCTCCCAAGCTGTGCACTGGGGGTGAG 292  
QY 1531 GCTGTGTCTACACACGCGGGGCTATCTTCCCTTGGGAGCCCAACGAGGAGAAACAGC 1590  
DB 293 GCTGTGTCTACACACGCGGGGCTATCTTCCCTTGGGAGCCCAACGAGGAGAAACAGC 352  
QY 1591 AACGATATTGCGCTGTGTCACCTCTCCAGTCCCTGCCCCCTACAG 1636  
DB 353 AACGATATTGCGCTGTGTCACCTCTCCAGTCCCTGCCCCCTACAG 398

RESULT 13  
US-09-820-002-16  
; Sequence 16, Application US/09820002  
; Patent No. 6482630  
; GENERAL INFORMATION:  
; APPLICANT: Gam, Weiniu  
; APPLICANT: Ye, Jane  
; APPLICANT: DiFrancesco, Valentina  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: C1001194  
; CURRENT APPLICATION NUMBER: US/09/820.002  
; CURRENT FILING DATE: 2001-03-29  
; NUMBER OF SEQ ID NOS: 16



SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-820-002-16

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.4e-38;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1446 GCGGAACCGGCTCTCTCCGAGTGTGCGAGTGTTCGCGGTGCGGAGCCGCTCTCC 1505  
DB 16 GCGGAACCGGCTCTCTCCGAGTGTGCGAGTGTTCGCGGTGCGGAGCCGCTCTCC 75  
QY 1506 CCACGCTGTGACGTGGGGGTGACAGCTGTGTCTACACAGGGGGCTATCTTCCCTTTCG 1565  
DB 76 CCACGCTGTGACGTGGGGGTGACAGCTGTGTCTACACAGGGGGCTATCTTCCCTTTCG 135  
QY 1566 GGAACCCCAACGCGAGGAACAAGATATTTCCCTGCTCCACTTCTCCAGTCCCT 1625  
DB 136 GGAACCCCAACGCGAGGAACAAGATATTTCCCTGCTCCACTTCTCCAGTCCCT 195  
QY 1626 GCCCTCTACAG 1636  
DB 196 GCCCTCTACAG 206

RESULT 14  
US-09-820-002-14  
; Sequence 14, Application US/09820002  
; Patent No. 6482630

GENERAL INFORMATION:  
; APPLICANT: Gan, Weiniu  
; APPLICANT: Ye, Jane  
; APPLICANT: DiFrancesco, Valentina  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01194  
; CURRENT APPLICATION NUMBER: US/09/820,002  
; CURRENT FILING DATE: 2001-03-29  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-820-002-14

Query Match  
Best Local Similarity 99.4%; Pred. No. 1.2e-32;  
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1278 AGACTGTGGCCGAGGAAGCTGCCCTGTGACCTGCTGGAGGCCGGAGACCAAGCTT 1337  
DB 320 AGACTGTGGCCGAGGAAGCTGCCCTGTGACCTGCTGGAGGCCGGAGACCAAGCTT 379  
QY 1338 GGGCCGCTGCGGTGGCAAGTCAAGCTTGGCTATGATGAGACACCTCTGTGGGGGATC 1397  
DB 380 GGGCCGCTGCGGTGGCAAGTCAAGCTTGGCTATGATGAGACACCTCTGTGTGGGGATC 439  
QY 1398 CTTGCTCTCCGGGGAAGTGGTGTGACAGCCGCCCACTGCTTCCGGAGCG 1448  
DB 440 CTTGCTCTCCGGGGAAGTGGTGTGACAGCCGCCCACTGCTTCCGGAGTG 490

RESULT 15  
US-09-949-016-45924  
; Sequence 45924, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45924  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-45924

Query Match  
Best Local Similarity 99.4%; Pred. No. 1.2e-32;  
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1278 AGACTGTGGCCGAGGAAGCTGCCCTGTGACCTGCTGGAGGCCGGAGACCAAGCTT 1337  
DB 324 AGACTGTGGCCGAGGAAGCTGCCCTGTGACCTGCTGGAGGCCGGAGACCAAGCTT 383  
QY 1338 GGGCCGCTGCGGTGGCAAGTCAAGCTTGGCTATGATGAGACACCTCTGTGGGGGATC 1397  
DB 384 GGGCCGCTGCGGTGGCAAGTCAAGCTTGGCTATGATGAGACACCTCTGTGTGGGGATC 443  
QY 1398 CTTGCTCTCCGGGGAAGTGGTGTGACAGCCGCCCACTGCTTCCGGAGCG 1448  
DB 444 CTTGCTCTCCGGGGAAGTGGTGTGACAGCCGCCCACTGCTTCCGGAGTG 494

Search completed: August 17, 2005, 10:50:38  
Job time : 559 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 01:11:50 ; Search time 1911 Seconds  
(without alignments)  
7319.910 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363  
Sequence: 1 tcgagccgcgttcctcagga.....taataagatggtttgatt 2363

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description         |
|------------|--------|-------------|--------|-------|---------------------|
| 1          | 2363   | 100.0       | 2363   | 6     | ABL68569 Kidney ca  |
| 2          | 2363   | 100.0       | 2363   | 6     | ADD40828 Human hep  |
| 3          | 2363   | 100.0       | 2363   | 6     | ADD40647 Human hep  |
| 4          | 2363   | 100.0       | 2363   | 6     | ADN97195 Gene #369  |
| 5          | 2363   | 100.0       | 2363   | 6     | ADBS5941 Human CDV  |
| 6          | 2363   | 100.0       | 2363   | 12    | ADLS6829 Human NOV  |
| 7          | 2363   | 100.0       | 2363   | 12    | ADOS9213 Human CDV  |
| 8          | 2303.4 | 97.5        | 2311   | 8     | ACCA451 Human dit   |
| 9          | 1638.6 | 69.3        | 1903   | 13    | ACNA1149 Human dia  |
| 10         | 1593.6 | 67.4        | 1783   | 4     | ADAD1167 Human hep  |
| 11         | 1593.6 | 67.4        | 1783   | 5     | AAS43103 CDNA enco  |
| 12         | 1593.6 | 67.4        | 1783   | 6     | ABO79532 Human hep  |
| 13         | 1593.6 | 67.4        | 1783   | 9     | ADAI7045 CDNA enco  |
| 14         | 1593.6 | 67.4        | 1783   | 10    | ADBS5353 Prostate   |
| 15         | 1593.6 | 67.4        | 1783   | 10    | ADD18428 Human pro  |
| 16         | 1593.6 | 67.4        | 1783   | 10    | ADII10403 Human cel |
| 17         | 1593.6 | 67.4        | 1783   | 11    | ADN39842 Cancer/an  |
| 18         | 1593.6 | 67.4        | 1783   | 12    | ADJ46927 Human tra  |
| 19         | 1593.6 | 67.4        | 1783   | 13    | ADRI8849 Human hep  |
| 20         | 1547.6 | 65.5        | 1811   | 13    | ACNA1153 Human dia  |

ALIGNMENTS

|    |        |      |       |    |          |                    |
|----|--------|------|-------|----|----------|--------------------|
| 21 | 1493.4 | 63.2 | 1758  | 13 | ACNA1152 | ACNA1152 Human dia |
| 22 | 1482   | 62.7 | 1482  | 6  | AAS62259 | AAS62259 CDNA sequ |
| 23 | 1446   | 61.2 | 1706  | 10 | ADJ79076 | ADJ79076 Human pro |
| 24 | 1405   | 59.5 | 1769  | 12 | ADJ57605 | ADJ57605 Human hep |
| 25 | 1390   | 58.8 | 1763  | 13 | ACNA1151 | ACNA1151 Human dia |
| 26 | 1349   | 57.1 | 1794  | 13 | ACNA1150 | ACNA1150 Human dia |
| 27 | 1269   | 53.7 | 1615  | 10 | ADBS2370 | ADBS2370 Human pro |
| 28 | 1269   | 53.7 | 1615  | 10 | ABX16007 | ABX16007 Human CDV |
| 29 | 1254   | 53.1 | 1254  | 12 | ADLS9245 | ADLS9245 Human CDV |
| 30 | 1098.6 | 46.5 | 1671  | 13 | ACNA1154 | ACNA1154 Human dia |
| 31 | 1097.8 | 46.5 | 7033  | 12 | ADDO7393 | ADDO7393 Modified  |
| 32 | 1097   | 46.4 | 12242 | 12 | ADDO7394 | ADDO7394 Modified  |
| 33 | 1031   | 43.6 | 1165  | 12 | ADDO7088 | ADDO7088 Human pro |
| 34 | 1029   | 43.5 | 1584  | 12 | ADDO6886 | ADDO6886 CDNA enco |
| 35 | 1023.6 | 43.3 | 1739  | 10 | ADBS9230 | ADBS9230 Toxicity- |
| 36 | 1023.6 | 43.3 | 1739  | 10 | ADBS2781 | ADBS2781 Primary r |
| 37 | 1008.2 | 42.7 | 1605  | 2  | AAI15134 | AAI15134 CDNA enco |
| 38 | 1008.2 | 42.7 | 1605  | 2  | AAZ31883 | AAZ31883 Mouse hep |
| 39 | 858.4  | 36.3 | 2175  | 11 | ADMO1874 | ADMO1874 Human CDV |
| 40 | 777.2  | 32.9 | 1192  | 12 | ADDO7395 | ADDO7395 Modified  |
| 41 | 449.4  | 19.0 | 644   | 13 | ADDO3467 | ADDO3467 Novel can |
| 42 | 402.8  | 17.0 | 493   | 9  | ACH45550 | ACH45550 Human foe |
| 43 | 391.2  | 16.6 | 614   | 6  | ABK30301 | ABK30301 Human G-P |
| 44 | 380.2  | 16.1 | 497   | 10 | ABX08968 | ABX08968 CDNA enco |
| 45 | 374.8  | 15.9 | 494   | 9  | ACH45563 | ACH45563 Human foe |

20030165839

PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 PA (AVAL-) AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 DR WPI; 2002-188264/24.  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1; SEQ ID NO 6906; 44pp; English.  
 XX  
 XX The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (1) of a signature gene set, where (1)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABU61664  
 CC to AB170110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (1) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
 CC tumour  
 CC  
 XX  
 SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2363; DB 6; Length 2363;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 181 TGAACCCAGGTAAGGACAAAGGGCCCCAGACTCAAGTTCCAGCCCTGAGGACAGGG 240  
 Qy 241 TTCCCTATCCCCCACCACCCAGCTTAATGCCCACTCTTAATAGAGGGTTCTGGGAC 300  
 Db 241 TTCCCTATCCCCCACCACCCAGCTTAATGCCCACTCTTAATAGAGGGTTCTGGGAC 300  
 Qy 301 TGAAGAGGGGACATAGAGTCTCCCCAGCACTAGTGTCTGCTGCTCTT 360  
 Db 301 TGAAGAGGGGACATAGAGTCTCCCCAGCACTAGTGTCTGCTGCTCTT 360  
 Qy 361 CAGACTAGCCGTTTGAAGCCCAAGTCTTTCTCCCAAGACCCAGAGTTCCAGCCCTAG 420  
 Db 361 CAGACTAGCCGTTTGAAGCCCAAGTCTTTCTCCCAAGACCCAGAGTTCCAGCCCTAG 420  
 Qy 421 GCCCTCTCTCTCATATCTAGGAGTCTGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480  
 Db 421 GCCCTCTCTCTCATATCTAGGAGTCTGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480  
 Qy 481 TGAATTCAGGTCCTCAGCTGTCTCTCTCTCAAAACGGAGTCTCAGTCCCTGCTCAC 540  
 Db 481 TGAATTCAGGTCCTCAGCTGTCTCTCTCTCAAAACGGAGTCTCAGTCCCTGCTCAC 540  
 Qy 541 CAGGCTCAGGCAATGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 600  
 Db 541 CAGGCTCAGGCAATGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 600  
 Qy 601 GACACTGACCCCAATCTTGAACCCAGGCCAATCGCTCGGTATCAAGGCGGCTGTGG 660  
 Db 601 GACACTGACCCCAATCTTGAACCCAGGCCAATCGCTCGGTATCAAGGCGGCTGTGG 660  
 Qy 661 CCAAGGCCAGTCCCTTACAGCTGCTGATGAGACGCTGGAGCTGGGGCGCCAGAGACT 720  
 Db 661 CCAAGGCCAGTCCCTTACAGCTGCTGATGAGACGCTGGAGCTGGGGCGCCAGAGACT 720  
 Qy 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGTCAATCTCTCAAGGTTCCACCC 780  
 Db 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGTCAATCTCTCAAGGTTCCACCC 780  
 Qy 781 TGGCCCGAGAGTCAAGCAGGGAATCTTAACAAGAGGCAATGAGGCGGAGAGAG 840  
 Db 781 TGGCCCGAGAGTCAAGCAGGGAATCTTAACAAGAGGCAATGAGGCGGAGAGAG 840  
 Qy 841 GGTGGCGAGCTGTCATCTGCTCCAGACCCAGGTGACACTCTGCGGGGACC 900  
 Db 841 GGTGGCGAGCTGTCATCTGCTCCAGACCCAGGTGACACTCTGCGGGGACC 900  
 Qy 901 CTGCTACTTGTGACAGGCAATCGGGGCGGCAATCTGAGGCTGTTCTCTCAGG 960  
 Db 901 CTGCTACTTGTGACAGGCAATCGGGGCGGCAATCTGAGGCTGTTCTCTCAGG 960  
 Qy 961 AGTGAACAGGAGCGGCTGTACCCAGTGCAGGTACAGCTGCGGAGACGCTGAGCTATGATC 1020  
 Db 961 AGTGAACAGGAGCGGCTGTACCCAGTGCAGGTACAGCTGCGGAGACGCTGAGCTATGATC 1020  
 Qy 1021 TTTGACAAAGACGGAAGGAGCTGAGCGGCTGCTCTGCGGCTCCAAAGCCAGGATA 1080  
 Db 1021 TTTGACAAAGACGGAAGGAGCTGAGCGGCTGCTCTGCGGCTCCAAAGCCAGGATA 1080  
 Qy 1081 GCCGAGTCAAGCTGCGAGAGATGGGCTTCTTAAGGCACTGACCTCCAGCTGAC 1140  
 Db 1081 GCCGAGTCAAGCTGCGAGAGATGGGCTTCTTAAGGCACTGACCTCCAGCTGAC 1140  
 Qy 1141 GTGCGAAGCGGGGCGCAATGCAAGTGGGCTTCTTGTGTGTGACGAGGGAGGCTG 1200  
 Db 1141 GTGCGAAGCGGGGCGCAATGCAAGTGGGCTTCTTGTGTGTGACGAGGGAGGCTG 1200  
 Qy 1201 CCCCAACCCAGAGGCTGTGAGAGTCAATCTCGTGTGTATGGCCCAAGGCGTTTC 1260  
 Db 1201 CCCCAACCCAGAGGCTGTGAGAGTCAATCTCGTGTGTATGGCCCAAGGCGTTTC 1260  
 Qy 1261 TTGGCGCATCTGCAAGACGTGTGGCGGAGGAAGTGCCTGCGAACCCGATGTGGGA 1320

```

Db 1261 TTGGCCGCATCTGCCAAGACTGTGCGCGAGGAAGCTGCCCGTGGAACCGCATCGTGGGA 1320
QY 1321 GGCCCGGACACCAAGCTTGGGCGGCTGGCCGCTGGAAGTCAAGCTTTCGCTATGATGAGCA 1380
Db 1321 GGCCCGGACACCAAGCTTGGGCGGCTGGCCGCTGGAAGTCAAGCTTTCGCTATGATGAGCA 1380
QY 1381 CACCTCTGTGGGGGATCCCTGCTCCGGGGGACTGGGGTGTGACAGAGCCGCCACAGCTTC 1440
Db 1381 CACCTCTGTGGGGGATCCCTGCTCCGGGGGACTGGGGTGTGACAGAGCCGCCACAGCTTC 1440
QY 1441 CCGAGCGGMAACGGGCTCTGTCCCGATGCGAGTGTTCGCGTGGCCGTCGAGCCAGCC 1500
Db 1441 CCGAGCGGMAACGGGCTCTGTCCCGATGCGAGTGTTCGCGTGGCCGTCGAGCCAGCC 1500
QY 1501 TCTCCCAACGGTCTGCACTGCGGGGTGCAAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Db 1501 TCTCCCAACGGTCTGCACTGCGGGGTGCAAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
QY 1561 TTTGGGGAGCCCAACAGGAGGAGAGAACAGACGATTTGGCCCTGTCCACCTCTCCAGT 1620
Db 1561 TTTGGGGAGCCCAACAGGAGGAGAGAACAGACGATTTGGCCCTGTCCACCTCTCCAGT 1620
QY 1621 CCCCTGCCCTCACAGAAATACATCCAGCTGTGTCTCCAGCTGCGGCGAGCCCTG 1680
Db 1621 CCCCTGCCCTCACAGAAATACATCCAGCTGTGTCTCCAGCTGCGGCGAGCCCTG 1680
QY 1681 GTGATGCGAAGATCTGTACCGTGAAGGGCTGGGGCAACAGCAGTACTATGCGCAAG 1740
Db 1681 GTGATGCGAAGATCTGTACCGTGAAGGGCTGGGGCAACAGCAGTACTATGCGCAAG 1740
QY 1741 GCGGGGGTACTCCAGAGGCTCGAGTCCCAATATACAAATGATGTGGAATGGCCCT 1800
Db 1741 GCGGGGGTACTCCAGAGGCTCGAGTCCCAATATACAAATGATGTGGAATGGCCCT 1800
QY 1801 GACTTCTATGGAACCAAGATCAAGGCCAAGTGTCTGTGCTGAGTACCCGAGGGTGC 1860
Db 1801 GACTTCTATGGAACCAAGATCAAGGCCAAGTGTCTGTGCTGAGTACCCGAGGGTGC 1860
QY 1861 ATTGATGCTGCGCAGGCGGACAGCGGTGTCTCTTGTGTGTGAGAGCAGATCTTCGG 1920
Db 1861 ATTGATGCTGCGCAGGCGGACAGCGGTGTCTCTTGTGTGTGAGAGCAGATCTTCGG 1920
QY 1921 ACGGCAGGTTGGCGGCTGTGTGGAATGTGTGAGTTGGGGCACTGTGCGCTGCGCCAG 1980
Db 1921 ACGGCAGGTTGGCGGCTGTGTGGAATGTGTGAGTTGGGGCACTGTGCGCTGCGCCAG 1980
QY 1981 AAGCCAGGCTGTACACCAAGTCAAGTACTTCGCGAGTGGATCTTCAGAGCCATAAG 2040
Db 1981 AAGCCAGGCTGTACACCAAGTCAAGTACTTCGCGAGTGGATCTTCAGAGCCATAAG 2040
QY 2041 ACTGACTCCGAAGCCAGCGGAGTGTGACCCAGCTCTGACCGGTGCTTTCGCTGCGCA 2100
Db 2041 ACTGACTCCGAAGCCAGCGGAGTGTGACCCAGCTCTGACCGGTGCTTTCGCTGCGCA 2100
QY 2101 GCGTCCAGAGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGTGGCGGAGATGGAGCT 2160
Db 2101 GCGTCCAGAGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGTGGCGGAGATGGAGCT 2160
QY 2161 TTTTCTTCTTGGGCGCGGTCTCACAGGTCCAGAGCAACCTCTCTCAGAGGCTCTCTCTC 2220
Db 2161 TTTTCTTCTTGGGCGCGGTCTCACAGGTCCAGAGCAACCTCTCTCAGAGGCTCTCTCTC 2220
QY 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCAACCACTCACCCTCTGACCCCAATGT 2280
Db 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCAACCACTCACCCTCTGACCCCAATGT 2280
QY 2281 AAAATATTTGTCTGCTGTGGAGACTCCCTGTCTAGTGTCCCTGTGATGATGGATCTCTTT 2340
Db 2281 AAAATATTTGTCTGCTGTGGAGACTCCCTGTCTAGTGTCCCTGTGATGATGGATCTCTTT 2340
QY 2341 AAAATATTAAGATGTTTGAAT 2363
Db 2341 AAAATATTAAGATGTTTGAAT 2363

```

```

RESULT 2
AAD40828
ID AAD40828 standard; DNA; 2363 BP.
XX
AC AAD40828;
XX
DT 30-OCT-2002 (first entry)
XX
DE Human hepsin DNA.
XX
KW Human; hepsin; antisense compound; antisense therapy; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 826..2079
FT /tag= a
FT /product= "Human hepsin protein"
XX
XX WO200250247-A2.
XX
XX 27-JUN-2002.
XX
XX 14-DEC-2001; 2001WO-US048341.
XX
XX 20-DEC-2000; 2000US-00742482.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Cowbert LM;
XX
XX WPI; 2002-519882/55.
XX
XX DR P-PSDB; AAE24850.
XX
XX PT Novel antisense compound targeted to nucleic acids encoding human hepsin,
XX useful for inhibiting the expression of hepsin in human cells or tissues,
XX and for treating humans having a disease associated with human hepsin.
XX
XX Example 13; Page 89-92; 100pp; English.
XX
CC The invention relates to antisense compounds, compositions and methods
CC for modulating the expression of hepsin. The compositions comprise
CC antisense compounds, particularly antisense oligonucleotides, targeted
CC to nucleic acids encoding hepsin. The antisense compound is useful for
CC inhibiting the expression of hepsin in human cells or tissues. It is also
CC useful for treating an animal having a disease or condition associated
CC with hepsin, by inhibiting expression of hepsin. It is useful for
CC diagnostics, therapeutics, prophylaxis and as research reagents and kits.
CC It is also used in antisense therapy. The present sequence is human
CC hepsin DNA
CC
SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
Query Match 100.0%; Score 2363; DB 6; Length 2363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGTGAAGGAGCTGCGCTAGC 60
Db 1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGTGAAGGAGCTGCGCTAGC 60
QY 61 AGGCCCAACGCAACCGCTGTGCTCCAGGCGCGGCTGTGCGGGGCAACATGCTCC 120
Db 61 AGGCCCAACGCAACCGCTGTGCTCCAGGCGCGGCTGTGCGGGGCAACATGCTCC 120
QY 121 TGGCCAGGCTTGAAGACTGACCGGACCGGCACTACTTGAAGGCTTCCGCCCACTGC 180
Db 121 TGGCCAGGCTTGAAGACTGACCGGACCGGCACTACTTGAAGGCTTCCGCCCACTGC 180
QY 181 TGGACCCGAGGATGAAGCAAGGGCCCGGAGCTGACAGTTCAGAGCCCTGAGGACAGGG 240
Db 181 TGGACCCGAGGATGAAGCAAGGGCCCGGAGCTGACAGTTCAGAGCCCTGAGGACAGGG 240

```

```
Db 181 TGAACCCCAAGGATTAAGACAAGAGGCCCCAGACTACAGTTCCAGCCCTGAGACAGAGG 240
Qy 241 TTCCCTCATCCCCCACCACCTTAATATGCCACTCTCTTAATAGAGGGTTCTTGGGACC 300
Db 241 TTCCCTCATCCCCCACCACCTTAATATGCCACTCTCTTAATAGAGGGTTCTTGGGACC 300
Qy 301 TGAAGAGGGGCACTATGACGTCTCCCAAGCACCTAGTGTCTGTCTCTCTCTT 360
Db 301 TGAAGAGGGGCACTATGACGTCTCCCAAGCACCTAGTGTCTGTCTCTCTCTT 360
Qy 361 CAGACTACAGCCGTTGAGACCCCAAGTCCTTTCTCCCAAGACCAGAGTTCCAGCCCTCAG 420
Db 361 CAGACTACAGCCGTTGAGACCCCAAGTCCTTTCTCCCAAGACCAGAGTTCCAGCCCTCAG 420
Qy 421 GCCCCCTCCCTCATATCAGTGGAGTCTGGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Db 421 GCCCCCTCCCTCATATCAGTGGAGTCTGGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Qy 481 TGATTTGAGTCTCAGCTGTCTCTCTCCCTCAAAACGGGATCTCAGTCCCTGCTCAC 540
Db 481 TGATTTGAGTCTCAGCTGTCTCTCTCCCTCAAAACGGGATCTCAGTCCCTGCTCAC 540
Qy 541 CAGGCTACAGGATGAGGAGTCCCAATCCCTGCAAAATCCAGAGGTCCTCCGCTGTGTCA 600
Db 541 CAGGCTACAGGATGAGGAGTCCCAATCCCTGCAAAATCCAGAGGTCCTCCGCTGTGTCA 600
Qy 601 GACACTACCCCATCTTTGAAACCAAGCCCAATCTGCGTCGTGATCAAGAGGTCTCTGG 660
Db 601 GACACTACCCCATCTTTGAAACCAAGCCCAATCTGCGTCGTGATCAAGAGGTCTCTGG 660
Qy 661 CCAAGGCCCAAGTCCCTACAGCTGCTGATGAGACGCTGAGGAGCTGGGGGCGCCAGAGACT 720
Db 661 CCAAGGCCCAAGTCCCTACAGCTGCTGATGAGACGCTGAGGAGCTGGGGGCGCCAGAGACT 720
Qy 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTTCATCTCTCAAGGTCCACC 780
Db 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTTCATCTCTCAAGGTCCACC 780
Qy 781 TGGCCCAAGAGGTGAGCCAGGGAATCATTAACAAGAGCAATGACATGAGGCAAGAGAG 840
Db 781 TGGCCCAAGAGGTGAGCCAGGGAATCATTAACAAGAGCAATGACATGAGGCAAGAGAG 840
Qy 841 GGTGGCCGAGCTGTGACATGTGCTCCAGACCCAAAGGTGAGAGCTCTCACTGCGGGACC 900
Db 841 GGTGGCCGAGCTGTGACATGTGCTCCAGACCCAAAGGTGAGAGCTCTCACTGCGGGACC 900
Qy 901 CTGCTACTTCTGACAGCCATCGGGGCGGCACTCTGGGCGCATTTGTGGCTTTCTCTCAGG 960
Db 901 CTGCTACTTCTGACAGCCATCGGGGCGGCACTCTGGGCGCATTTGTGGCTTTCTCTCAGG 960
Qy 961 AGTACCAAGAGCCGCTGTATCCAGTGCAGTTCAGTCTGAGACGCTCCGGCTCATGGTC 1020
Db 961 AGTACCAAGAGCCGCTGTATCCAGTGCAGTTCAGTCTGAGACGCTCCGGCTCATGGTC 1020
Qy 1021 TTTTGAACAAGACGAAGGGACGTGCGGCTGTGCTGCTCTGCGCTCCAAAGCCAGAGGTA 1080
Db 1021 TTTTGAACAAGACGAAGGGACGTGCGGCTGTGCTGCTCTGCGCTCCAAAGCCAGAGGTA 1080
Qy 1081 GCCGGAAGTCAAGCTCGAAGAGATGAGGCTTCTCTCAAGGCACTGACCCACTCGAAGTGGAC 1140
Db 1081 GCCGGAAGTCAAGCTCGAAGAGATGAGGCTTCTCTCAAGGCACTGACCCACTCGAAGTGGAC 1140
Qy 1141 GTTGGAAAGGAGGCGGCAATGAGACGTGGGCTTCTCTGTGTGAGACGAAGGAGAGCTG 1200
Db 1141 GTTGGAAAGGAGGCGGCAATGAGACGTGGGCTTCTCTGTGTGAGACGAAGGAGAGCTG 1200
Qy 1201 CCCCCAAGCCAGAGAGCTGTGAGAGTCACTCGTGTGTGATTTGCCAGAGGCGGTTTTC 1260
Db 1201 CCCCCAAGCCAGAGAGCTGTGAGAGTCACTCGTGTGTGATTTGCCAGAGGCGGTTTTC 1260
Qy 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCGGAGAGCTGCCGTGGAGCCGATGTGGGA 1320
Db 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCGGAGAGCTGCCGTGGAGCCGATGTGGGA 1320
Qy 1321 GGGCGGGACACAAGCTTGGGCGGGTGGCCGAGGCAAGTCAAGCTTGCTATGATGAGACA 1380
Db 1321 GGGCGGGACACAAGCTTGGGCGGGTGGCCGAGGCAAGTCAAGCTTGCTATGATGAGACA 1380
Qy 1381 CACCTGTGTGGGAGATCCCTGCTCTCGGGGAACTGGGTGTGAAGCCGCGCCACTGTCTTC 1440
Db 1381 CACCTGTGTGGGAGATCCCTGCTCTCGGGGAACTGGGTGTGAAGCCGCGCCACTGTCTTC 1440
Qy 1441 CCGGAGCGGAAACGGGGTCTGTCCCATGCGAGGTGTTTCCGGGTGCGGTCAGAGCC 1500
Db 1441 CCGGAGCGGAAACGGGGTCTGTCCCATGCGAGGTGTTTCCGGGTGCGGTCAGAGCC 1500
Qy 1501 TCTCCCAAGGTCGACGCTGGGGGTGACAGCTGTGGTCTACCAAGGGGAGTACTTCTCC 1560
Db 1501 TCTCCCAAGGTCGACGCTGGGGGTGACAGCTGTGGTCTACCAAGGGGAGTACTTCTCC 1560
Qy 1561 TTTTGGGACCCCAACAGCGAGAGAAACAGCAAGATATGCTCCGTGTCACTCTCCAGT 1620
Db 1561 TTTTGGGACCCCAACAGCGAGAGAAACAGCAAGATATGCTCCGTGTCACTCTCCAGT 1620
Qy 1621 CCGCTGCGCTCACAGAAATCATTCAGCTGTGTGCTCCAGCTGCGGCGCAGGCGCTG 1680
Db 1621 CCGCTGCGCTCACAGAAATCATTCAGCTGTGTGCTCCAGCTGCGGCGCAGGCGCTG 1680
Qy 1681 GTGGATGGCAAGATCTGTACCGTGAAGGCTGGGGGCAACGAGATCTATGGCCAAAG 1740
Db 1681 GTGGATGGCAAGATCTGTACCGTGAAGGCTGGGGGCAACGAGATCTATGGCCAAAG 1740
Qy 1741 GCGGGGTAAGTCCAGAGAGGCTGAGTCCCATATACAGATGATGTCTGCAATGGGCT 1800
Db 1741 GCGGGGTAAGTCCAGAGAGGCTGAGTCCCATATACAGATGATGTCTGCAATGGGCT 1800
Qy 1801 GACTTCTATGGAACCAAGATCAAGCCCAATATGTTCTGTCTGTACCCCGAGGGTGGC 1860
Db 1801 GACTTCTATGGAACCAAGATCAAGCCCAATATGTTCTGTCTGTACCCCGAGGGTGGC 1860
Qy 1861 ATTGATGCTGCGCAGAGGCGACAGCGGTGTCCTCTTGTGTGTGAGACACACTCTCGG 1920
Db 1861 ATTGATGCTGCGCAGAGGCGACAGCGGTGTCCTCTTGTGTGTGAGACACACTCTCGG 1920
Qy 1921 ACGCCAGTGGCGGCTGTGTGAGTGTGAGTGTGGGCACTGAGCTGTGCTCGGCCAG 1980
Db 1921 ACGCCAGTGGCGGCTGTGTGAGTGTGAGTGTGGGCACTGAGCTGTGCTCGGCCAG 1980
Qy 1981 AAGCCAGGCGTCTACACAAAGTCAAGTCACTTCCGGGAGTGAATCTTCCAGGCCATPAAAG 2040
Db 1981 AAGCCAGGCGTCTACACAAAGTCAAGTCACTTCCGGGAGTGAATCTTCCAGGCCATPAAAG 2040
Qy 2041 ACTCACTCCGAGAGCGAGCGGATGTGACCCAGCTGTGACCGGAGCTTCTCGCTGGCA 2100
Db 2041 ACTCACTCCGAGAGCGAGCGGATGTGACCCAGCTGTGACCGGAGCTTCTCGCTGGCA 2100
Qy 2101 GCTTCCAGGGCCGAGGTGATCCCGGTGTGGAGTCCACGCTGGGCGGAGATGGAGCT 2160
Db 2101 GCTTCCAGGGCCGAGGTGATCCCGGTGTGGAGTCCACGCTGGGCGGAGATGGAGCT 2160
Qy 2161 TTTTCTTCTTGGGCGGGTCAAGGTCCAAAGACACCTTCTCCAGGGTCTCTCTTTC 2220
Db 2161 TTTTCTTCTTGGGCGGGTCAAGGTCCAAAGACACCTTCTCCAGGGTCTCTCTTTC 2220
Qy 2221 CACAGTGGCGGGGCGACTCAGCCCGGAGACCAACCACTCAACCTCTCAACCCCACTGT 2280
Db 2221 CACAGTGGCGGGGCGACTCAGCCCGGAGACCAACCACTCAACCTCTCAACCCCACTGT 2280
Qy 2281 AAATATTTGTTCTGCTGTCTGGAATCTCTGTCTAAGTGTCCCTGATGATGAGATGCTCTT 2340
Db 2281 AAATATTTGTTCTGCTGTCTGGAATCTCTGTCTAAGTGTCCCTGATGATGAGATGCTCTT 2340
Qy 2341 AAATATTAAGATGATTTTGAAT 2363
Db 2341 AAATATTAAGATGATTTTGAAT 2363
```



RESULT 3  
AAD40647 standard; DNA; 2363 BP.  
XX AAD40647;  
AC  
XX  
30-OCT-2002 (first entry)  
DT  
XX  
Human hepsin DNA.  
DE  
XX  
Human; hepsin; inflammation; tumour; gene therapy; cytostatic; gene; ds.  
KM  
XX  
Homo sapiens.  
OS  
XX  
Key Location/Qualifiers  
FH 826..2079  
FT /\*tag= a  
PT CDS /product= "Human hepsin"  
XX  
XX  
XX W0200250248-A2.  
XX  
XX 27-JUN-2002.  
XX  
XX 14-DEC-2001; 2001WO-US048431.  
XX  
XX 20-DEC-2000; 2000US-00742703.  
XX  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX (ABBO) ABBOTT LAB.  
XX  
XX Marcotte PA, Cowbert LM;  
XX  
XX WPI; 2002-519883/55.  
XX P-PSDB; AAE25075.  
XX  
XX New antisense oligonucleotides that modulate (particularly inhibit) human  
PT hepsin, useful for treating a disease or condition associated with the  
PT expression of hepsin, e.g. inflammation or tumor growth.  
XX  
XX  
XX Example 13; Page 88-91; 101pp; English.  
XX  
XX The invention relates to an antisense compound 8-30 nucleobases in length  
CC targeted to a nucleic acid molecule encoding human hepsin. The antisense  
CC compound specifically hybridises with and inhibits the expression of  
CC human hepsin. The antisense compound or the pharmaceutical composition is  
CC useful for treating animals and humans having a disease or condition  
CC associated with the expression of hepsin, e.g. inflammation or tumour  
CC growth. The antisense compounds are useful also for diagnostics,  
CC prophylaxis (e.g. to prevent or delay infection, inflammation or tumour  
CC formation) or as research reagents and kits. The method is useful for  
CC modulating, specifically inhibiting the expression of hepsin which may be  
CC used in research, e.g. to distinguish between functions of various members  
CC of a biological pathway. The invention is used in gene therapy. The  
CC present sequence is human hepsin DNA  
XX  
XX  
XX Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 2363; DB 6; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGGTGAGGAGCCTGCGCTAGC 60  
DB 1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGGTGAGGAGCCTGCGCTAGC 60  
QY 61 AGGCGCCACGCGACCGGCTGCTCCAGAGCGCGCGCGCTGCTGCGGGGCCACATGCTCC 120  
DB 61 AGGCGCCACGCGACCGGCTGCTCCAGAGCGCGCGCGCTGCTGCGGGGCCACATGCTCC 120  
QY 121 TGCCCAAGCCTGAGAGTGAACCCGACCTACCTCGAGGCTCGCGCCCACTGC 180  
DB 121 TGCCCAAGCCTGAGAGTGAACCCGACCTACCTCGAGGCTCGCGCCCACTGC 180

QY 181 TGAACCCCAAGGTTAAGGACAGAGGCCCCAGACTCAAGTTCAGCCCTGAGGACAGGG 240  
DB 181 TGAACCCCAAGGTTAAGGACAGAGGCCCCAGACTCAAGTTCAGCCCTGAGGACAGGG 240  
QY 241 TTCCCTCATCCCCCAAGCCTTAATGAGGAGGTTCTTGAGGAGC 300  
DB 241 TTCCCTCATCCCCCAAGCCTTAATGAGGAGGTTCTTGAGGAGC 300  
QY 301 TGAAGAGGAGGACATGAGGCTTCCCAAGACCTTAGTGTGTTCTGCTCTTCTT 360  
DB 301 TGAAGAGGAGGACATGAGGCTTCCCAAGACCTTAGTGTGTTCTGCTCTTCTT 360  
QY 361 CAGACTGAGGCTTGAACCCCAAGTCTTCTTCCCAAGACCCAGGAGTTCCAGGCTCAG 420  
DB 361 CAGACTGAGGCTTGAACCCCAAGTCTTCTTCCCAAGACCCAGGAGTTCCAGGCTCAG 420  
QY 421 GCCCTCTCTCTCACTAGGAGTCTGAGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480  
DB 421 GCCCTCTCTCTCACTAGGAGTCTGAGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480  
QY 481 TGATTTGAGGCTCTGAGCTGTCTCTCTCAAAACCGGATCTCAGTCCCTGCTCAC 540  
DB 481 TGATTTGAGGCTCTGAGCTGTCTCTCTCAAAACCGGATCTCAGTCCCTGCTCAC 540  
QY 541 CAGGCTCAGGCTAGGGGGTCCCATCCCTGCAAAATCCAGGCTCCCCCGCTGCTGCTCA 600  
DB 541 CAGGCTCAGGCTAGGGGGTCCCATCCCTGCAAAATCCAGGCTCCCCCGCTGCTGCTCA 600  
QY 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTGCTGATCAAGGCTGCTGAG 660  
DB 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTGCTGATCAAGGCTGCTGAG 660  
QY 661 CCAAGGCCAAGTCCCTCAAGCTGCTGATGAGACGCTGAGACTGAGGAGCGCCAGACT 720  
DB 661 CCAAGGCCAAGTCCCTCAAGCTGCTGATGAGACGCTGAGACTGAGGAGCGCCAGACT 720  
QY 721 GGGCTGAGGCTGGGCTCCCCAGGCGCTGCTCCCTGCTCAATCTCTCAAGTCCCAACC 780  
DB 721 GGGCTGAGGCTGGGCTCCCCAGGCGCTGCTCCCTGCTCAATCTCTCAAGTCCCAACC 780  
QY 781 TGAGCCAGAGGCTCAGGACGAGGATCATTAACAAGAGGAGTGAATGAGGAGAGAG 840  
DB 781 TGAGCCAGAGGCTCAGGACGAGGATCATTAACAAGAGGAGTGAATGAGGAGAGAG 840  
QY 841 GGTGAGCCGAGCTGTGCATCTGCTCAGACCCAGGTGAGCTCTACTGCGGAGACC 900  
DB 841 GGTGAGCCGAGCTGTGCATCTGCTCAGACCCAGGTGAGCTCTACTGCGGAGACC 900  
QY 901 CTGCTACTTGTGACAGGATCGGGGCGGACATCTGCGGACATTTGTGCTGCTCTAGG 960  
DB 901 CTGCTACTTGTGACAGGATCGGGGCGGACATCTGCGGACATTTGTGCTGCTCTAGG 960  
QY 961 AGTGACAGGAGCGGCTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020  
DB 961 AGTGACAGGAGCGGCTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020  
QY 1021 TTTGACAAAGCGAGAGGAGGAGTGTGCTGCTCTGCGCTTCAACGCCAGGAGTA 1080  
DB 1021 TTTGACAAAGCGAGAGGAGGAGTGTGCTGCTCTGCGCTTCAACGCCAGGAGTA 1080  
QY 1081 GCGGAGTCAAGCTGCGAGAGAGTGTGCTTCTTCAAGGACCTCAAGCTGAGAGT 1140  
DB 1081 GCGGAGTCAAGCTGCGAGAGAGTGTGCTTCTTCAAGGACCTCAAGCTGAGAGT 1140  
QY 1141 GTGCGAAGCGGAGGCGCAATGAGAGTGTGCTTCTTGTGTGAGCGAGGAGGCTG 1200  
DB 1141 GTGCGAAGCGGAGGCGCAATGAGAGTGTGCTTCTTGTGTGAGCGAGGAGGCTG 1200  
QY 1201 CCCCACACCCAGAGGCTGTGAGAGTCACTCCGTGTGTATGCCCCAGAGGCGTTTC 1260  
DB 1201 CCCCACACCCAGAGGCTGTGAGAGTCACTCCGTGTGTATGCCCCAGAGGCGTTTC 1260

```
OY 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGCGAGAAAGCTGCCGTGGACCCGCGATCGTGGGA 1320
    |||
DB 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGCGAGAAAGCTGCCGTGGACCCGCGATCGTGGGA 1320
OY 1321 GGGCGGGGACACAGCTTGGGCGGGTGGCCGTGGCAAGTCAGCTTCCTGCTATGATGAGACA 1380
    |||
DB 1321 GGGCGGGGACACAGCTTGGGCGGGTGGCCGTGGCAAGTCAGCTTCCTGCTATGATGAGACA 1380
OY 1381 CACCTCTGTGGGGGATCCCTGCTCTCCGGGAGCTGGGTGTGACAGCCGCCCATCTGCTTC 1440
    |||
DB 1381 CACCTCTGTGGGGGATCCCTGCTCTCCGGGAGCTGGGTGTGACAGCCGCCCATCTGCTTC 1440
OY 1441 CCGGAGCGGAAACCGGGTCTGTGCCGATGCGAGGTGTTGGCGGGTCCGTGGCCAGAGCC 1500
    |||
DB 1441 CCGGAGCGGAAACCGGGTCTGTGCCGATGCGAGGTGTTGGCGGGTCCGTGGCCAGAGCC 1500
OY 1501 TCTCCCAACGGTCTGACAGCTGGGGGTGCAAGCTGTGGTCTACACAGGGGGCTATCTTCC 1560
    |||
DB 1501 TCTCCCAACGGTCTGACAGCTGGGGGTGCAAGCTGTGGTCTACACAGGGGGCTATCTTCC 1560
OY 1561 TTTTGGGAGCCCAACAGGAGAGAAACAGCAAGATATGGCCCTGTCTCACTCTCCAGT 1620
    |||
DB 1561 TTTTGGGAGCCCAACAGGAGAGAAACAGCAAGATATGGCCCTGTCTCACTCTCCAGT 1620
OY 1621 CCCCTGCCCTCAGAGAAATACATCAGGCTGTGTGCTCCAGGCTGCGGCGAGAGCCCTG 1680
    |||
DB 1621 CCCCTGCCCTCAGAGAAATACATCAGGCTGTGTGCTCCAGGCTGCGGCGAGAGCCCTG 1680
OY 1681 GTGGATGGCAAGATCTGTACCGGTGACGGGCTGGGGCAACAGCAGTACTATGAGCCAAAG 1740
    |||
DB 1681 GTGGATGGCAAGATCTGTACCGGTGACGGGCTGGGGCAACAGCAGTACTATGAGCCAAAG 1740
OY 1741 GCGCGGGTATCTCCAGAGGCTCGAGTCCCCCATATCAGCAATGATGCTGAAATGGGCT 1800
    |||
DB 1741 GCGCGGGTATCTCCAGAGGCTCGAGTCCCCCATATCAGCAATGATGCTGAAATGGGCT 1800
OY 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCCGAGGGTGGC 1860
    |||
DB 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCCGAGGGTGGC 1860
OY 1861 ATTGATGCTTGGCCAGGGCGAGCGGGTGTCTTTGTGTGTGAGAGACAGATCTTCGG 1920
    |||
DB 1861 ATTGATGCTTGGCCAGGGCGAGCGGGTGTCTTTGTGTGTGAGAGACAGATCTTCGG 1920
OY 1921 ACGCCAGGTTGGCGGCTGTGTGTGAGATTTGGGCACTGGCTGTGGCTTGGCCAG 1980
    |||
DB 1921 ACGCCAGGTTGGCGGCTGTGTGTGAGATTTGGGCACTGGCTGTGGCTTGGCCAG 1980
OY 1981 AAGCCAGGCTGTACACCAAGTCAAGTCTTCGGGAGTGGATCTTCCAGGCGATAAG 2040
    |||
DB 1981 AAGCCAGGCTGTACACCAAGTCAAGTCTTCCGGAGTGGATCTTCCAGGCGATAAG 2040
OY 2041 ACTGACTTCGAAGCAGCGGAGTGTGACCCAGCTCTGACCGGTGGCTTCTGCTGGCA 2100
    |||
DB 2041 ACTGACTTCGAAGCAGCGGAGTGTGACCCAGCTCTGACCGGTGGCTTCTGCTGGCA 2100
OY 2101 GCCCTCAGGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGGGCGGAGATGGGAGT 2160
    |||
DB 2101 GCCCTCAGGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGGGCGGAGATGGGAGT 2160
OY 2161 TTTTCTTCTTGGGCGCGGTCCACAGGTTCAGAGACACCTCTCCAGAGGTCTCTCTTC 2220
    |||
DB 2161 TTTTCTTCTTGGGCGCGGTCCACAGGTTCAGAGTTCAGAGACACCTCTCCAGAGGTCTCTCTTC 2220
OY 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCACTCAACCTCAGCCCTCTGACCCCATGT 2280
    |||
DB 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCACTCAACCTCAGCCCTCTGACCCCATGT 2280
OY 2281 AAAAATTTGTTCTGTCTGTGTGGAGATCTCTGTCTAGGTGCCCCGTGATGATGGATGCTCTTT 2340
    |||
DB 2281 AAAAATTTGTTCTGTCTGTGTGGAGATCTCTGTCTAGGTGCCCCGTGATGATGGATGCTCTTT 2340
OY 2341 AAATATATTAAGATGTTTGAAT 2363
```

```
DB 2341 AAATATATTAAGATGTTTGAAT 2363
|||||
RESULT 4
ID ABR97195 standard; DNA; 2363 BP.
XX
AC ABR97195;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3693 used to diagnose liver cancer.
XX
KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KM metastatic liver tumour; cytostatic; expression profile; disease state;
KM disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN W0200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 3693; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABR93503-ABR97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 2363; DB 6; Length 2363;
Best local similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGGCCACAGTGAAGGAGCTGGCTAGC 60
    |||
DB 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGGCCACAGTGAAGGAGCTGGCTAGC 60
OY 61 AGGCCCCACCCAGCGCTTGTCTTCAGAGCCGCCGCTGTGGGCGCACATGCTCC 120
    |||
DB 61 AGGCCCCACCCAGCGCTTGTCTTCAGAGCCGCCGCTGTGGGCGCACATGCTCC 120
OY 121 TGCCAGGCTTGAAGACTGACCCGAGCCCGGACATCTGAGGCTTCGCCCCCACTGCG 180
    |||
DB 121 TGCCAGGCTTGAAGACTGACCCGAGCCCGGACATCTGAGGCTTCGCCCCCACTGCG 180
```

181 TGGACCCAGGGGTAAAGGCAAGGGGCCCCAGACTGACAGTTCCAGGCTGAGAGAGAGGG 240  
Db TGGACCCAGGGGTAAAGGCAAGGGGCCCCAGACTGACAGTTCCAGGCTGAGAGAGAGGG 240  
Qy TTCCCTCATCCCGCCAGCCAGCTTAATGAGGAGGTTCTTGAGGAGC 300  
Db TTCCCTCATCCCGCCAGCCAGCTTAATGAGGAGGTTCTTGAGGAGC 300  
Qy TGAAGAGGGGGGCACTATGACGTCCTCCAGAGCACTAGATGTTGTCCTGCTTCCCTT 360  
Db TGAAGAGGGGGGCACTATGACGTCCTCCAGAGCACTAGATGTTGTCCTGCTTCCCTT 360  
Qy CAGACTGAGCGGTGAGAGCCAGAGTCTTCTCCAGAGCACTAGAGGTTCCAGGCTTCCAG 420  
Db CAGACTGAGCGGTGAGAGCCAGAGTCTTCTCCAGAGCACTAGAGGTTCCAGGCTTCCAG 420  
Qy GCGCCCTCTCTCATCTAGAGGAGTCTTGAGCCCAATTCCTCTTCCAGAGCTTA 480  
Db GCGCCCTCTCTCATCTAGAGGAGTCTTGAGCCCAATTCCTCTTCCAGAGCTTA 480  
Qy TGAATTCAGGTCCTCAGCTGTCCTCTCCAAACCGGAGTCTTCAATCCCTGCTCCAG 540  
Db TGAATTCAGGTCCTCAGCTGTCCTCTCCAAACCGGAGTCTTCAATCCCTGCTCCAG 540  
Qy CAGGCTCAGGATGAGGGGTCCCAATCCCTGCAAAATCCAGAGTCCCGGCTGCTGAGCA 600  
Db CAGGCTCAGGATGAGGGGTCCCAATCCCTGCAAAATCCAGAGTCCCGGCTGCTGAGCA 600  
Qy GACACTGACCCCATCTTGAACCCAGCCCAATCTGCGTCCGATGACAGGAGTCTGAG 660  
Db GACACTGACCCCATCTTGAACCCAGCCCAATCTGCGTCCGATGACAGGAGTCTGAG 660  
Qy CCAAAGGCCAGTCCCTACAGCCTGCTGATGAGAGCCTGAGGAGTGGGGGCGCCAGAGCT 720  
Db CCAAAGGCCAGTCCCTACAGCCTGCTGATGAGAGCCTGAGGAGTGGGGGCGCCAGAGCT 720  
Qy GGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCTGATCTCTCAAGAGTCCCAACC 780  
Db GGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCTGATCTCTCAAGAGTCCCAACC 780  
Qy TGGCCAGAGAGGTGAGCCAGGGAATCATTAACAAGAGGAGTGAATGGCCAGAGAGAG 840  
Db TGGCCAGAGAGGTGAGCCAGGGAATCATTAACAAGAGGAGTGAATGGCCAGAGAGAG 840  
Qy GGTGGCCGAGTGTGCAATGCTGCTCCAGAGCCAAAGTGGAGCTTCACTGCGGAGCC 900  
Db GGTGGCCGAGTGTGCAATGCTGCTCCAGAGCCAAAGTGGAGCTTCACTGCGGAGCC 900  
Qy CTGCTACTTCTGACAGCCATGAGGGGAGCTTGGGCAATTTGAGCTGTTCTCTCAGG 960  
Db CTGCTACTTCTGACAGCCATGAGGGGAGCTTGGGCAATTTGAGCTGTTCTCTCAGG 960  
Qy AGTACCAAGAGCGGTGATCCAGTGAAGTGAAGTCTGAGGAGCGCTCGGCTCATGCTC 1020  
Db AGTACCAAGAGCGGTGATCCAGTGAAGTGAAGTCTGAGGAGCGCTCGGCTCATGCTC 1020  
Qy TTTTGAACAAGAGGAGCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Db TTTTGAACAAGAGGAGCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Qy GCGGAGCTCAGCTGAGAGAGATGAGGCTTCTCAGGAGCACTGAGCCACTCGAGCTGAGC 1140  
Db GCGGAGCTCAGCTGAGAGAGATGAGGCTTCTCAGGAGCACTGAGCCACTCGAGCTGAGC 1140  
Qy GTGCGAAGCGGCGGCGCAATGAGCACTGCGGCTTCTCTGCTGAGAGCAAGAGGAGGCTG 1200  
Db GTGCGAAGCGGCGGCGCAATGAGCACTGCGGCTTCTCTGCTGAGAGCAAGAGGAGGCTG 1200  
Qy CCGCAACACCAAGAGCTGCTGAGAGTCACTCTCGTGTGATGAGCCCAAGGCGGCTTTC 1260  
Db CCGCAACACCAAGAGCTGCTGAGAGTCACTCTCGTGTGATGAGCCCAAGGCGGCTTTC 1260

Qy TTGGCCGCATCTGCAAGACTGTGGCCGAGAGAGCTGCGGCTGAGAGAGAGGAG 1320  
Db TTGGCCGCATCTGCAAGACTGTGGCCGAGAGAGCTGCGGCTGAGAGAGAGGAG 1320  
Qy GCGCGGAGCAACAGCTTGGGCGGCTGAGGAGTCAAGCTTCTGCTATGATGAGAGCA 1380  
Db GCGCGGAGCAACAGCTTGGGCGGCTGAGGAGTCAAGCTTCTGCTATGATGAGAGCA 1380  
Qy CACTCTGTGGGGATCCCTGCTCTCGGGGAGCTGGGTGCTGACAGCCGCCCACTGCTTC 1440  
Db CACTCTGTGGGGATCCCTGCTCTCGGGGAGCTGGGTGCTGACAGCCGCCCACTGCTTC 1440  
Qy CCGAGCGGAACCGGCTCTGCTCCGATGAGGAGGTTTGCGGAGCGGTGGCCAGGCGC 1500  
Db CCGAGCGGAACCGGCTCTGCTCCGATGAGGAGGTTTGCGGAGCGGTGGCCAGGCGC 1500  
Qy TCTCCCAAGGCTGACAGTGGGGGAGCTGAGTCTACACAGGGGGCTATCTTCC 1560  
Db TCTCCCAAGGCTGACAGTGGGGGAGCTGAGTCTACACAGGGGGCTATCTTCC 1560  
Qy TTTGGGAGCCCAACAGGAGAGAGCAAGAGATATGCTGCTGCTCAGCT 1620  
Db TTTGGGAGCCCAACAGGAGAGAGCAAGAGATATGCTGCTGCTCAGCT 1620  
Qy TTTGGGAGCCCAACAGGAGAGAGCAAGAGATATGCTGCTGCTCAGCT 1620  
Db TTTGGGAGCCCAACAGGAGAGAGCAAGAGATATGCTGCTGCTCAGCT 1620  
Qy GTGATGAGCAAGATCTGACCTGAGCGGCTGAGGAGCAACAGCACTATGAGCCAG 1740  
Db GTGATGAGCAAGATCTGACCTGAGCGGCTGAGGAGCAACAGCACTATGAGCCAG 1740  
Qy GTGATGAGCAAGATCTGACCTGAGCGGCTGAGGAGCAACAGCACTATGAGCCAG 1740  
Db GTGATGAGCAAGATCTGACCTGAGCGGCTGAGGAGCAACAGCACTATGAGCCAG 1740  
Qy GCGGGGCTACTCCAGAGGCTGAGTCCCAATACAGATGATGTCTGCAATGAGCGCT 1800  
Db GCGGGGCTACTCCAGAGGCTGAGTCCCAATACAGATGATGTCTGCAATGAGCGCT 1800  
Qy GCGGGGCTACTCCAGAGGCTGAGTCCCAATACAGATGATGTCTGCAATGAGCGCT 1800  
Db GCGGGGCTACTCCAGAGGCTGAGTCCCAATACAGATGATGTCTGCAATGAGCGCT 1800  
Qy GACTTCTATGAGAAACAGATCAACAGTCAAGTCAAGTCTGCTGCTGCTGCTGCTGCTGCT 1860  
Db GACTTCTATGAGAAACAGATCAACAGTCAAGTCAAGTCTGCTGCTGCTGCTGCTGCTGCT 1860  
Qy ATTGATGCTGCTGCAAGGCGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920  
Db ATTGATGCTGCTGCAAGGCGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920  
Qy ACGCCAGCTGGGCGGCTGTGTGAGTGTGAGTGGGAGCACTGCTGCTGCTGCTGCTGCTGCT 1980  
Db ACGCCAGCTGGGCGGCTGTGTGAGTGTGAGTGGGAGCACTGCTGCTGCTGCTGCTGCTGCT 1980  
Qy AAGCCAGGCGCTTACCAAAAGTGAAGTCTTCCGGGAGTGGATCTTCCAGGCCATTAAG 2040  
Db AAGCCAGGCGCTTACCAAAAGTGAAGTCTTCCGGGAGTGGATCTTCCAGGCCATTAAG 2040  
Qy ACTCACTCCGAAGCGAGGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGAT 2100  
Db ACTCACTCCGAAGCGAGGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGAT 2100  
Qy GCTCTCAAGGCGCGAGTGTATCCCGGTGTGTGATGATCAAGCTGAGGAGTGTGAGTGTGAT 2160  
Db GCTCTCAAGGCGCGAGTGTATCCCGGTGTGTGATGATCAAGCTGAGGAGTGTGAGTGTGAT 2160  
Qy TTTTCTTCTTGTGGGCGGCTGCAAGGTTCMAAGGACACCTTCTCCAGGCTCTCTCTTC 2220  
Db TTTTCTTCTTGTGGGCGGCTGCAAGGTTCMAAGGACACCTTCTCCAGGCTCTCTCTTC 2220  
Qy CACAGTGGCGGCGCACTCAGCCCGAGAGCAACCAACTCAACCTCTCTGAGCCCACTGT 2280  
Db CACAGTGGCGGCGCACTCAGCCCGAGAGCAACCAACTCAACCTCTCTGAGCCCACTGT 2280  
Qy AAATATGTTCTGCTGTCTGAGTCTCTGCTGATGAGTCCCTGATGATGAGGATCTCTT 2340  
Db AAATATGTTCTGCTGTCTGAGTCTCTGCTGATGAGTCCCTGATGATGAGGATCTCTT 2340  
Qy AAATATGTTCTGCTGTCTGAGTCTCTGCTGATGAGTCCCTGATGATGAGGATCTCTT 2340  
Db AAATATGTTCTGCTGTCTGAGTCTCTGCTGATGAGTCCCTGATGATGAGGATCTCTT 2340



|    |      |  |      |
|----|------|--|------|
| Dp | 1643 | GGGCTGGGCTGGGCTCCCCAGAGCCCTGCTCCCGTCACTCTCCACAGCTCCACCC        | 1584 |
| Oy | 791  | TGGCCAGAGGTCAGCCAGGGAAATCATTAACAAGGCAGTGACATGCGCAGAGAG         | 840  |
| Dp | 1583 | TGGCCAGAGGTCAGCCAGGGAAATCATTAACAAGGCAGTGACATGCGCAGAGAG         | 1524 |
| Oy | 841  | GGTGGCCGGAATGTGTCATGTCTGCTCCAGACCAGAGTGCGAGCTCTACTGCGGGACC     | 900  |
| Dp | 1523 | GGTGGCCGGAATGTGTCATGTCTGCTCCAGACCAGAGTGCGAGCTCTACTGCGGGACC     | 1464 |
| Oy | 901  | CTGCTACTTCTGACAGCCATCGGGGCGGACATCTCGGGGCATTGTGGGCTGTCTCTCAGG   | 960  |
| Dp | 1463 | CTGCTACTTCTGACAGCCATCGGGGCGGACATCTCGGGGCATTGTGGGCTGTCTCTCAGG   | 1404 |
| Oy | 961  | AGTGACAGAGCCGCTGTACCCAGTGCAGGTCAAGCTGTGGAGACGTGCGCTCATGCTC     | 1020 |
| Dp | 1403 | AGTGACAGAGCCGCTGTACCCAGTGCAGGTCAAGCTGTGGAGACGTGCGCTCATGCTC     | 1344 |
| Oy | 1021 | TTTTCACAAAGAGGAAAGGACGTGGCGGCTGTGCTCTCTCGGCTCCAAAGCAGGTA       | 1080 |
| Dp | 1343 | TTTTCACAAAGAGGAAAGGACGTGGCGGCTGTGCTCTCTCGGCTCCAAAGCAGGTA       | 1284 |
| Oy | 1081 | GCCGAGCTCAGCTGGAGAGATGGGGCTTCCTCAAGGGACGACCCACTCGAGCTGGAC      | 1140 |
| Dp | 1283 | GCCGAGCTCAGCTGGAGAGATGGGGCTTCCTCAAGGGACGACCCACTCGAGCTGGAC      | 1224 |
| Oy | 1141 | GTGCGAAGCGGGCGGCGCAATGCGACGTGCGGGCTTCTTCTGTGTGACAGAGGGAGGCTG   | 1200 |
| Dp | 1223 | GTGCGAAGCGGGCGGCGCAATGCGACGTGCGGGCTTCTTCTGTGTGACAGAGGGAGGCTG   | 1164 |
| Oy | 1201 | CCCCACACCAAGAGGCTGCTGAGGTCATCTCCGTGTGATTTGCCCCAGAGGCCGTTTC     | 1260 |
| Dp | 1163 | CCCCACACCAAGAGGCTGCTGAGGTCATCTCCGTGTGATTTGCCCCAGAGGCCGTTTC     | 1104 |
| Oy | 1261 | TTGGCCCGGCATCTGCGAAGACTGTGAGCCGCGAAGAGCTGCCCGTGGACCCGATGTGGGA  | 1320 |
| Dp | 1103 | TTGGCCCGGCATCTGCGAAGACTGTGAGCCGCGAAGAGCTGCCCGTGGACCCGATGTGGGA  | 1044 |
| Oy | 1321 | GGCGGGGACACAGAGCTTGGGCGGGTGGCCGTGGCAAGTCAGACCTTCCGTATGATGAGACA | 1380 |
| Dp | 1043 | GGCGGGGACACAGAGCTTGGGCGGGTGGCCGTGGCAAGTCAGACCTTCCGTATGATGAGACA | 984  |
| Oy | 1381 | CACCTCTGTGGGGGATGCCCTGCTCTCCGGGGAGTTGGGTGTGACAGCCGCCACTGCTTC   | 1440 |
| Dp | 983  | CACCTCTGTGGGGGATGCCCTGCTCTCCGGGGAGTTGGGTGTGACAGCCGCCACTGCTTC   | 924  |
| Oy | 1441 | CCGGAGCGGAACCGGGTCTCTGTCCCGATGCGAGTGTGGCCGCTGCCGTCAGAGCC       | 1500 |
| Dp | 923  | CCGGAGCGGAACCGGGTCTCTGTCCCGATGCGAGTGTGGCCGCTGCCGTCAGAGCC       | 864  |
| Oy | 1501 | TCTCCCAACGGTCTGACAGCTGGGGGTGAGGCTGTGGTCTACACAGGGGGCTATCTTCCC   | 1560 |
| Dp | 863  | TCTCCCAACGGTCTGACAGCTGGGGGTGAGGCTGTGGTCTACACAGGGGGCTATCTTCCC   | 804  |
| Oy | 1561 | TTTTCGGGACCCCAACAGCGAGAGAAACGACATATTGCCCTGTGTCTCACTCTCCAGT     | 1620 |
| Dp | 803  | TTTTCGGGACCCCAACAGCGAGAGAAACGACATATTGCCCTGTGTCTCACTCTCCAGT     | 744  |
| Oy | 1621 | CCCTGCCCCCTCAAGAAATACATCAGCTGTGTGCTCCAGCTGCCGCGCAGAGCCCTG      | 1680 |
| Dp | 743  | CCCTGCCCCCTCAAGAAATACATCAGCTGTGTGCTCCAGCTGCCGCGCAGAGCCCTG      | 684  |
| Oy | 1681 | GTGATGCGCAAGATCTGTACGCTACCGGGGCTGGGGCAACAGCGATCTATTGGCGAACG    | 1740 |
| Dp | 683  | GTGATGCGCAAGATCTGTACGCTACCGGGGCTGGGGCAACAGCGATCTATTGGCGAACG    | 624  |
| Oy | 1741 | GCCGGGATCTCCAGAGAGGCTCGAGTCCCACTAAATCAGCAATGATGTCTCGAATGGCGCT  | 1800 |
| Dp | 623  | GCCGGGATCTCCAGAGAGGCTCGAGTCCCACTAAATCAGCAATGATGTCTCGAATGGCGCT  | 564  |
| Oy | 1801 | GACTTCTATTGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCGAGGGTGGC       | 1860 |

|            |   |                 |  |      |
|------------|---|-----------------|--|------|
| Db         |   | 563             | GACCTTCTATGSAAGAACAGATCGAAGCCCAAGATGTTCTGTGTGGCTCACCAGAGGGTGGC | 504  |
| Oy         |   | 1861            | ATTGATGCTCTGCCAGGGCGACAGCGGTGTCTCCCTTGTGTGTGAAGAACAAGATCTCTCGG | 1920 |
| Db         |   | 503             | ATTGATGCTCTGCCAGGGCGACAGCGGTGTCTCCCTTGTGTGTGAAGAACAAGATCTCTCGG | 444  |
| Oy         |   | 1921            | AAGCCACGTTGGCCGCTGTGTGGCATTTGTAGTTGGGGCACTGGCTGTGCCCTGGCCAG    | 1980 |
| Db         |   | 443             | ACGCCACTTGGCCGCTGTGTGGCATTTGTAGTTGGGGCACTGGCTGTGCCCTGGCCAG     | 384  |
| Oy         |   | 1981            | AAGCAGGCGCTCTACACCAAGATCAGTAGCTTCGGGAGTGAATCTTCCAGGCCATAAAG    | 2040 |
| Db         |   | 383             | AAGCAGGCGCTCTACACCAAGATCAGTAGCTTCGGGAGTGAATCTTCCAGGCCATAAAG    | 324  |
| Oy         |   | 2041            | ACTCATCTCCGAAGCCAGCGGAGTGTGACCCCAAGCTTGTGACCGAGTGTCTCTCGCTCGCA | 2100 |
| Db         |   | 323             | ACTCATCTCCGAAGCCAGCGGAGTGTGACCCCAAGCTTGTGACCGAGTGTCTCTCGCTCGCA | 264  |
| Oy         |   | 2101            | GCCTCCAGAGGGCCGAGGTGATCCCCTGGTGTGGGATTCACGCTGGGCCGAGATGGGACGT  | 2160 |
| Db         |   | 263             | GCCTCCAGAGGGCCGAGGTGATCCCCTGGTGTGGGATTCACGCTGGGCCGAGATGGGACGT  | 204  |
| Oy         |   | 2161            | TTTTCTTCTTGGGCCCCGCTCCACAGATCCAAAGACACCTTCCCTCCAGAGGCTCTCTTC   | 2220 |
| Db         |   | 203             | TTTTCTTCTTGGGCCCCGCTCCACAGATCCAAAGACACCTTCCCTCCAGAGGCTCTCTTC   | 144  |
| Oy         |   | 2221            | CACAGTGGCGGGCCCACTCAGCCCCCGAGACCAACCACTCACCCCTCTGACCCCAATGT    | 2280 |
| Db         |   | 143             | CACAGTGGCGGGCCCACTCAGCCCCCGAGACCAACCACTCACCCCTCTGACCCCAATGT    | 84   |
| Oy         |   | 2281            | AAATATTGTTGTGCTGTCTGGGACTCCTGTCTAGGTGCCCCCTGATGATGGGATGCTTTT   | 2340 |
| Db         |   | 83              | AAATATTGTTGTGCTGTCTGGGACTCCTGTCTAGGTGCCCCCTGATGATGGGATGCTTTT   | 24   |
| Oy         |   | 2341            | AAATPATPAAGATGTTTGAATT   | 2363 |
| Db         |   | 23              | AAATPATPAAGATGTTTGAATT   | 1    |
| <hr/>      |   |                 |  |      |
| RESULT 6   |   |                 |  |      |
| ADL56829/C |   |                 |  |      |
| ID         | ADL56829  | standard; cDNA; | 2363 BP.   |      |
| XX         |   |                 |  |      |
| AC         | ADL56829;   |                 |  |      |
| XX         |   |                 |  |      |
| DT         | 20-MAY-2004   | (first entry)   |  |      |
| XX         |   |                 |  |      |
| DE         | Human NOX cDNA #8.  |                 |  |      |
| XX         |   |                 |  |      |
| KM         | Human; NOX; gene; ss; angiogenic-associated disorder; cancer;             |                 |  |      |
| KW         | cardiovascular disease; psoriasis; wound healing; stroke; cardiomyopathy; |                 |  |      |
| KM         | atherosclerosis; cell signal processing disorder;                         |                 |  |      |
| KM         | metabolic pathway modulation disorder; diabetes;                          |                 |  |      |
| KM         | hyperproliferative disease; cirrhosis; keloid; psoriasis; osteoarthritis; |                 |  |      |
| KW         | haemorrhage; ischaemic heart disease; renal disease; thrombosis;          |                 |  |      |
| KM         | hypertension; hypothyroidism; autoimmune disorder; multiple sclerosis;    |                 |  |      |
| KM         | systemic lupus erythematosus; rheumatoid arthritis;                       |                 |  |      |
| KM         | autoimmune pulmonary inflammation; Guillain-Barre syndrome;               |                 |  |      |
| KM         | autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;     |                 |  |      |
| KW         | asthma; periodontal disease; Alzheimer's disease; Parkinson's disease;    |                 |  |      |
| KM         | Huntington's disease; amyotrophic lateral sclerosis; shy-drager syndrome. |                 |  |      |
| XX         |   |                 |  |      |
| OS         | Homo sapiens.   |                 |  |      |
| PM         | US2003215449-A1.  |                 |  |      |
| XX         |   |                 |  |      |
| PD         | 20-NOV-2003.  |                 |  |      |
| XX         |   |                 |  |      |
| PF         | 15-MAR-2002; 2002US-00099322.   |                 |  |      |
| XX         |   |                 |  |      |
| PR         | 11-JAN-2001; 2001US-0261013P.   |                 |  |      |
| PR         | 11-JAN-2001; 2001US-0261014P.   |                 |  |      |
| PR         | 11-JAN-2001; 2001US-0261018P.   |                 |  |      |

PR 11-JAN-2001; 2001US-0261026P.  
PR 11-JAN-2001; 2001US-0261029P.  
PR 23-MAR-2001; 2001US-0278152P.  
PR 17-AUG-2001; 2001US-0313170P.  
PR 10-SEP-2001; 2001US-0318410P.  
PR 11-JAN-2002; 2002US-00044564.

XX  
PA (MEZE/) MEZES P D.  
PA (RAST/) RASTELLI L.  
PA (HERR/) HERRMANN J L.  
PA (MACD/) MACDOUGALL J R.  
PA (ZHON/) ZHONG H.  
PA (CASW/) CASMAN S J.  
PA (BOLD/) BOLDOG F L.  
PA (SHIM/) SHIMKETS R A.  
PA (GORM/) GORMAN L.  
PA (ETSE/) EISEN A J.  
PA (SPAD/) SPADERNA S K.  
PA (VERN/) VERNET C A M.  
PA (BERG/) BERGHS C.  
PA (SPYT/) SPYTEK K A.  
PA (DIP1/) DIPIRPO V A.  
PA (ZERH/) ZERHUSEN B D.  
PA (PEYM/) PEYMAN J A.  
PA (ELLE/) ELLERMAN K.  
PA (STON/) STONE D J.  
PA (GROS/) GROSSE W M.  
PA (ALSO/) ALSOBROOK J P.  
PA (LEPL/) LEPLLEY D M.  
PA (RIEG/) RIEGER D K.  
PA (BURG/) BURGESS C E.  
PA (EDIN/) EDINGER S R.  
PA (VOSS/) VOSS E Z.  
PA (MILL/) MILLER C E.

XX  
PI Mezes PD, Rastelli L, Herrmann JL, Macdougall JR, Zhong H;  
PI Caman SJ, Boldog FL, Shimkets RA, Gorman L, Eisen AJ, Spaderna SK;  
PI Vernet CM, Bergs C, Spytek KA, Dipippo VA, Zernusen BD, Peyman JA;  
PI Ellerman K, Stone DJ, Grose WM, Alsobrook JP, Lepley DM, Rieger DK;  
PI Burgess CE, Edinger SR, Voss EZ, Miller CE;  
XX  
DR WPI; 2004-130990/13.  
DR P-PSDB; ADL56830.

PT Novel isolated SECL and/or NOXY polypeptide, useful for treating cancer,  
PT cardiovascular disease, psoriasis, wound healing, and stroke.

XX  
PS Claim 24; SEQ ID NO 39; 315pp; English.

CC The invention relates to human SECX and NOXY polypeptides and the  
CC polynucleotides encoding them. The invention also relates to antibodies  
CC that bind immunospecifically to the polypeptides. The sequences are  
CC useful for treating or preventing angiogenic-associated disorders,  
CC cancer, cardiovascular disease, psoriasis, wound healing, stroke,  
CC cardiomyopathy, atherosclerosis, cell signal processing disorders,  
CC metabolic pathway modulation disorders, diabetes, hyperproliferative  
CC diseases, cirrhosis, keloids, psoriasis, osteoarthritis, atherosclerotic  
CC plaque formation, haemorrhage, ischaemic heart disease, renal disease,  
CC thrombosis, hypertension, hypothyroidism, autoimmune disorders, multiple  
CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune  
CC pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis,  
CC myasthenia gravis, graft-versus-host disease, asthma, periodontal  
CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis and shy-drager syndrome. This sequence  
CC represents a human NOXY polynucleotide of the invention.

XX  
SQ Sequence 2363 BP; 456 A; 696 C; 808 G; 403 T; 0 U; 0 Other;

Query Match 100.0%; Score 2363; DB 12; Length 2363;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGAGCCGCTTTCAGGAGACCTTACCTGAGGCGCCACAGGTGAGGCGAGCTGAGCTTAC 60

DB 2363 TCGAGCCGCTTTCAGGAGACCTTACCTGAGGCGCCACAGGTGAGGCGAGCTGAGCTTAC 2304  
QY AGGCCCCAGCCACCGGCTCTGCTCCAGGCGCGGCTGTGGGGGCCACCATGCTCC 120  
DB AGGCCCCAGCCACCGGCTCTGCTCCAGGCGCGGCTGTGGGGGCCACCATGCTCC 2244  
QY TGCCAGGCGCTGAGAGCTGAGCCGAGCCGCGGACCTACCTGAGGCTCGGCCCACTGC 180  
DB TGCCAGGCGCTGAGAGCTGAGCCGAGCCGCGGACCTACCTGAGGCTCGGCCCACTGC 2184  
QY TGAGCCCGAGGTAAGGACAAGGGCCCCAGACTCAGATTCCAGCCCTGAGACAGGG 240  
DB TGAGCCCGAGGTAAGGACAAGGGCCCCAGACTCAGATTCCAGCCCTGAGACAGGG 2124  
QY TTCCCTATCCCCCAGCCAGCTTAATGCCACCTCTTAATAGGGGTTCTGGGGACC 300  
DB TTCCCTATCCCCCAGCCAGCTTAATGCCACCTCTTAATAGGGGTTCTGGGGACC 2064  
QY TGAAGAGGGGGCACTATGAGGTCTCCCAAGACCTAGTGTCTGTCTGCTTCTT 360  
DB TGAAGAGGGGGCACTATGAGGTCTCCCAAGACCTAGTGTCTGTCTGCTTCTT 2004  
QY CAGACTCAGCGGTTGAGACCCAGTCTTTCCTCCAGACCTAGGTTCTGCTTCTT 420  
DB CAGACTCAGCGGTTGAGACCCAGTCTTTCCTCCAGACCTAGGTTCTGCTTCTT 2004  
QY GCGCCCTCTCTCTATCTAGGAGTCTGCCCCCAATATTCCTCTTCCCAAGACTTA 480  
DB GCGCCCTCTCTCTATCTAGGAGTCTGCCCCCAATATTCCTCTTCCCAAGACTTA 1943  
QY GCGCCCTCTCTCTATCTAGGAGTCTGCCCCCAATATTCCTCTTCCCAAGACTTA 1884  
DB GCGCCCTCTCTCTATCTAGGAGTCTGCCCCCAATATTCCTCTTCCCAAGACTTA 1883  
QY TGATTTAGGTTCTCAGCTGTCTCTCTCTCAAAACGGGATCTCAGTCCCTGCTCAC 540  
DB TGATTTAGGTTCTCAGCTGTCTCTCTCTCAAAACGGGATCTCAGTCCCTGCTCAC 1824  
QY CAGGCTCAGGCTAGGGGGTCCCATCCCTGCAATTCAGGCGTCCCGCTGTGTCA 600  
DB CAGGCTCAGGCTAGGGGGTCCCATCCCTGCAATTCAGGCGTCCCGCTGTGTCA 1823  
QY GACACTACCCCATCTCTTGAACCAAGCCAAATCTGCTCTCTATACAGGCTGCTTG 660  
DB GACACTACCCCATCTCTTGAACCAAGCCAAATCTGCTCTCTATACAGGCTGCTTG 1763  
QY GACACTACCCCATCTCTTGAACCAAGCCAAATCTGCTCTCTATACAGGCTGCTTG 1704  
DB GACACTACCCCATCTCTTGAACCAAGCCAAATCTGCTCTCTATACAGGCTGCTTG 661  
QY CCAAGGCGCAGTCCCTTACAGCTGCTGAGTGAAGCGCTGAGACTGAGGAGCT 720  
DB CCAAGGCGCAGTCCCTTACAGCTGCTGAGTGAAGCGCTGAGACTGAGGAGCT 1703  
QY GGGCTGGGCTGGGCTCCCGAGGCTGCTCCCGTCCATCTCTCAAGGTCCACCC 780  
DB GGGCTGGGCTGGGCTCCCGAGGCTGCTCCCGTCCATCTCTCAAGGTCCACCC 1643  
QY TGCGCCAGAGGTTCAGGCAAGGAATCTTAACAAGAGCGATGACATGAGGAGAG 840  
DB TGCGCCAGAGGTTCAGGCAAGGAATCTTAACAAGAGCGATGACATGAGGAGAG 1583  
QY TGCGCCAGAGGTTCAGGCAAGGAATCTTAACAAGAGCGATGACATGAGGAGAG 1524  
DB TGCGCCAGAGGTTCAGGCAAGGAATCTTAACAAGAGCGATGACATGAGGAGAG 841  
QY GGTGCGGAGCTGTCATCTGCTCCAGACCAAGGTGAGCTCTCACTGCGGGACC 900  
DB GGTGCGGAGCTGTCATCTGCTCCAGACCAAGGTGAGCTCTCACTGCGGGACC 1523  
QY GTGGCGGAGCTGTCATCTGCTCCAGACCAAGGTGAGCTCTCACTGCGGGACC 1464  
DB GTGGCGGAGCTGTCATCTGCTCCAGACCAAGGTGAGCTCTCACTGCGGGACC 901  
QY CTGCTACTTTCAGACGCAATCGGGGCGGATCTCTGGGCAATGTGGCTTCTTCAG 960  
DB CTGCTACTTTCAGACGCAATCGGGGCGGATCTCTGGGCAATGTGGCTTCTTCAG 1463  
QY CTGCTACTTTCAGACGCAATCGGGGCGGATCTCTGGGCAATGTGGCTTCTTCAG 1404  
DB CTGCTACTTTCAGACGCAATCGGGGCGGATCTCTGGGCAATGTGGCTTCTTCAG 961  
QY AGTAGACAGAGAGCGCTGTATCCAGTGCAGTCACTCTGCGGATATGTC 1020  
DB AGTAGACAGAGAGCGCTGTATCCAGTGCAGTCACTCTGCGGATATGTC 1403  
QY AGTAGACAGAGAGCGCTGTATCCAGTGCAGTCACTCTGCGGATATGTC 1344  
DB AGTAGACAGAGAGCGCTGTATCCAGTGCAGTCACTCTGCGGATATGTC 1021  
QY TTGACAG 1080  
DB TTGACAG 1343  
QY TTGACAG 1284  
DB TTGACAG 1081  
QY GCGGACTCAGCTGCGAGAGATGAGGCTTCTTCAAGGCACTGACCTCGAGCTGAC 1140  
DB GCGGACTCAGCTGCGAGAGATGAGGCTTCTTCAAGGCACTGACCTCGAGCTGAC



Db 1283 GCCGAGCTCAGCTGCGAGAGATGAGGCTTCTCAGGAGCAGTACCACCTCCGAGCTGAGC 1224  
Qy 1141 GTGGCAAGCGCGGGGCCCAATGGCACGTGGGGCTTCTTGTGTGTGAGAGAGGGAGGCTG 1200  
Db 1223 GTGCAAGCGCGGGGCCCAATGGCACGTGGGGCTTCTTGTGTGTGAGAGAGGGAGGCTG 1164  
Qy 1201 CCCACACCCAGAGGCTGTGGAGGTCACTCCGTGTGTGATGTGCCCCAGAGGCCGTTTC 1260  
Db 1163 CCCACACCCAGAGGCTGTGGAGGTCACTCCGTGTGTGATGTGCCCCAGAGGCCGTTTC 1104  
Qy 1261 TTGGCCGCATCTGCGAAGACTGTGAGCGGAGAGTCCCGTGGAGCGCATCTGTGGA 1320  
Db 1103 TTGGCCGCATCTGCGAAGACTGTGAGCGGAGAGTCCCGTGGAGCGCATCTGTGGA 1044  
Qy 1321 GGCCGGGACACCACTTGGGCGGTGGCCGTGGCAAGTCAAGCTTCCCTTGTGTGAGCA 1380  
Db 1043 GGCCGGGACACCACTTGGGCGGTGGCCGTGGCAAGTCAAGCTTCCCTTGTGTGAGCA 984  
Qy 1381 CACCTGTGTGGGGGATCCCTGCTCCGGGGGACTGGGGTGTGAGACGGCCGCCACTGCTTC 1440  
Db 983 CACCTGTGTGGGGGATCCCTGCTCCGGGGGACTGGGGTGTGAGACGGCCGCCACTGCTTC 924  
Qy 1441 CCGAGCGGAACCGGGTCTGTCCCGATGCGAGTGTGGCCGAGTCCGCTGGCCAGGCC 1500  
Db 923 CCGAGCGGAACCGGGTCTGTCCCGATGCGAGTGTGGCCGAGTGTGGCCGAGGCCAGGCC 864  
Qy 1501 TCTCCCAACGGTCTGACACTGGGGGTGACGGTGTGTCTACACAGGGAGCTATTTTCC 1560  
Db 863 TCTCCCAACGGTCTGACACTGGGGGTGACGGTGTGTCTACACAGGGAGCTATTTTCC 804  
Qy 1561 TTTCGGGACCCCAACAGGAGGAGAGACAGACGATATTTGCCCTGTGTCTCCTCCCACT 1620  
Db 803 TTTCGGGACCCCAACAGGAGGAGAGACAGACGATATTTGCCCTGTGTCTCCTCCCACT 744  
Qy 1621 CCCCTGCCCTCACAGAAATACACGAGCTGTGTCTCCAGCTGCCGCGACAGGCCCTG 1680  
Db 743 CCCCTGCCCTCACAGAAATACACGAGCTGTGTCTCCAGCTGCCGCGACAGGCCCTG 684  
Qy 1681 GTGATGCGAAGATTTGATCCGTGACGGGCTGGGGCAACAGCAGTACTATGGCCACAG 1740  
Db 683 GTGATGCGAAGATTTGATCCGTGACGGGCTGGGGCAACAGCAGTACTATGGCCACAG 624  
Qy 1741 GCCGGGATCTCCAGAGGAGCTCGAGTCCCAATACAGCATGATGTGAAATGGCGCT 1800  
Db 623 GCCGGGATCTCCAGAGGAGCTCGAGTCCCAATACAGCATGATGTGAAATGGCGCT 564  
Qy 1801 GACTTCTATGGAACCAAGATCAAGCCAAAGTGTCTGTGCTGCTACCCGAGGGTGGC 1860  
Db 563 GACTTCTATGGAACCAAGATCAAGCCAAAGTGTCTGTGCTGCTACCCGAGGGTGGC 504  
Qy 1861 ATTGATGCTGCGAGGCGAGACAGCGGTGTCTTGTGTGTGAGAGCAGATCTTGGG 1920  
Db 503 ATTGATGCTGCGAGGCGAGACAGCGGTGTCTTGTGTGTGAGAGCAGATCTTGGG 444  
Qy 1921 ACGGCACTTGGCGGCTGTGTGCAATTGTGAGTTGGGGCACTGGGTGGCCCTGCGCCAG 1980  
Db 443 ACGGCACTTGGCGGCTGTGTGCAATTGTGAGTTGGGGCACTGGGTGGCCCTGCGCCAG 384  
Qy 1981 AAGCAGGAGCTCTACACCAAGATCAGTATTCGCGAGGTGGAATTTTCAGAGCCATAAG 2040  
Db 383 AAGCAGGAGCTCTACACCAAGATCAGTATTCGCGAGGTGGAATTTTCAGAGCCATAAG 324  
Qy 2041 ACTCACTCGAAGCCAGGCGAGTGTGACCCAGCTGTGACCGGTGGCTTCTGCTGCGCA 2100  
Db 323 ACTCACTCGAAGCCAGGCGAGTGTGACCCAGCTGTGACCGGTGGCTTCTGCTGCGCA 264  
Qy 2101 GCGTCAGAGGCGCGAGGTGATCCGCGTGGTGGAGTCCAGCGTGGCGGAGAGATGGAGCT 2160  
Db 263 GCGTCAGAGGCGCGAGGTGATCCGCGTGGTGGAGTCCAGCGTGGCGGAGAGATGGAGCT 204  
Qy 2161 TTTTCTTCTTGGGCGCGGTGCCAGAGTCCAGAGACACCTCTCTCAGAGGTCTCTCTTC 2220  
Db 203 TTTTCTTCTTGGGCGCGGTGCCAGAGTCCAGAGACACCTCTCTCTCAGAGGTCTCTCTTC 144

Qy 2221 CACAGTGGGCGGCGCCACTCAGCCCGGAGACCAACCAACTGACCTCTGACCCCGCATGT 2280  
Db 143 CACAGTGGGCGGCGCCACTCAGCCCGGAGACCAACCAACTGACCTCTGACCCCGCATGT 84  
Qy 2281 AAATATTGTTGCTGTGTGGAGTCTGTCTAGAGTGGCCCTGATGATGGATGCTTTT 2340  
Db 83 AAATATTGTTGCTGTGTGGAGTCTGTCTAGAGTGGCCCTGATGATGGATGCTTTT 24  
Qy 2341 AAATATTAAAGATGGTTTGATT 2363  
Db 23 AAATATTAAAGATGGTTTGATT 1  
RESULT 7  
ID ADO39213 standard; cDNA; 2363 BP.  
XX ADO39213;  
XX 15-JUL-2004 (first entry)  
DE Human cDNA encoding novel protein NOV8.  
XX  
KW Human; ss; gene; NOVX; cardiomyopathy; atherosclerosis; hypertension;  
KW congenital heart defect; aortic stenosis; atrial septal defect;  
KW atrioventricular canal defect; ductus arteriosus; pulmonary stenosis;  
KW subaortic stenosis; ventricular septal defect; valve disease;  
KW tuberosus sclerosis; scleroderma; obesity; transplantation;  
KW congenital adrenal hyperplasia; prostate cancer; neoplasm;  
KW adenocarcinoma; lymphoma; uterus cancer; haemophilia; hypercoagulation;  
KW idiopathic thrombocytopenic purpura; immunodeficiency;  
KW giant versus host disease; AIDS; bronchial asthma; Crohn's disease;  
KW multiple sclerosis; angiogenic-associated disorder; psoriasis;  
wound healing; stroke.  
XX OS Homo sapiens.  
XX US2004018196-A1.  
XX PD 29-JAN-2004.  
XX  
XX 11-JAN-2002; 2002US-00044564.  
XX PE 11-JAN-2001; 2001US-0261013P.  
XX PR 11-JAN-2001; 2001US-0261014P.  
XX PR 11-JAN-2001; 2001US-0261018P.  
XX PR 11-JAN-2001; 2001US-0261026P.  
XX PR 11-JAN-2001; 2001US-0261029P.  
XX PR 17-AUG-2001; 2001US-0313170P.  
XX PR 10-SEP-2001; 2001US-0318410P.  
XX  
XX (MEZE/) MEZES P D.  
XX (RAST/) RASTELLI L.  
XX (HERR/) HERRMANN J L.  
XX (MACD/) MACDOUGALL J R.  
XX (ZHON/) ZHONG H.  
XX (CASW/) CASMAN S J.  
XX (BOLD/) BOLDOG F L.  
XX (SHIM/) SHIMKETS R A.  
XX (GORM/) GORMAN L.  
XX (EISE/) EISEN A J.  
XX (SPAD/) SPADERNA S K.  
XX (VERN/) VERNET C A M.  
XX (BERG/) BERGHS C.  
XX (SPYT/) SPYTEK K A.  
XX (DIPI/) DIPRO V A.  
XX (ZERH/) ZERHUSEN B D.  
XX (PEYM/) PEYMAN J A.  
XX (ELIE/) ELLERMAN K.  
XX (STON/) STONE D J.  
XX (GROS/) GROSSE W M.  
XX (ALSO/) ALSOBROOK J P.

PA (LEPL/) LEPLLEY D M.  
 PA (RIEG/) RIEGER D K.  
 PA (BURG/) BURGESS C R.  
 PA (EDIN/) EDINGER S R.  
 PI Meese PD, Rastelli L, Herrmann JL, MacDougall JR, Zhong H,  
 PI Caman SJ, Boldog FL, Shukets RA, Gorman L, Eisen AJ, Spederna SK;  
 PI Vernet CAM, Berghs C, Sptek KA, Dipippo VA, Zernusen BD, Peyman JA;  
 PI Ellerman K, Stone DJ, Grose WM, Alsbrook JP, Lepley DM, Rieger DK;  
 PI Burgess CE, Edinger SR;  
 XX WPI: 2004-121988/12.  
 DR P-PSDB; ADO39214.  
 XX  
 PT New isolated SECK polypeptides and polynucleotides, useful for treating  
 PT or preventing, e.g. cardiomyopathy, atherosclerosis, obesity, prostate  
 PT cancer, hemophilia, graft versus host disease, Crohn's disease, multiple  
 PT sclerosis or psoriasis.  
 PS  
 PS Claim 23; SEQ ID NO 39; 306bp; English.  
 XX  
 CC The invention relates to a method of treating or delaying the onset of an  
 CC angiogenic-associated disorder comprising administering an antibody to  
 CC the polypeptide of SECK1. The cDNAs and proteins for SECK1-SECK12 are  
 CC disclosed as new. Also included are an isolated NOVX (NOV1-NOV8)  
 CC polypeptide (or its mature form or variant), the encoding nucleic acid  
 CC for the NOVX protein described above, a vector comprising the nucleic  
 CC acid molecule, a cell comprising the vector, an antibody that binds  
 CC immunospecifically to NOVX, a method of treating or delaying the onset of  
 CC an angiogenic-associated disorder, a method for determining the presence  
 CC of or predisposition to a disease associated with altered levels of SECK1  
 CC (or altered levels of the nucleic acid of SECK1 in a first mammalian  
 CC subject), a method for determining the presence of or predisposition to a  
 CC disease associated with altered levels of SECK1-SECK12, a method of  
 CC treating a pathological state in a mammal, a method of treating or  
 CC delaying the onset of a disorder, a method for determining the presence  
 CC or amount of the polypeptide or nucleic acid, a method of identifying an  
 CC agent that binds to a polypeptide, a method for identifying an agent that  
 CC modulates the expression or activity of the polypeptide, a method for  
 CC modulating the activity of the polypeptide, a method of treating or  
 CC preventing a SECK-associated or NOVX disorder, a pharmaceutical  
 CC composition (comprising NOVX or SECK, the nucleic acid molecule or the  
 CC antibody and a pharmaceutical carrier), a kit comprising (in one or more  
 CC containers) the pharmaceutical composition and a method for determining  
 CC the presence of or predisposition to a disease associated with altered  
 CC levels of the polypeptide or the nucleic acid in a first mammalian  
 CC subject. The SECK/NOVX polypeptides and polynucleotides are useful for  
 CC cardiomyopathy, atherosclerosis, hypertension, congenital heart defects,  
 CC aortic stenosis, atrial septal defect, atrioventricular canal defect,  
 CC ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular  
 CC septal defect, valve disease, tuberosus sclerosis, scleroderma, obesity,  
 CC transplantation, congenital adrenal hyperplasia, prostate cancer,  
 CC neoplasm, adenocarcinoma, lymphoma, uterine cancer, haemophilia,  
 CC hypercoagulation, idiopathic thrombocytopenic purpura,  
 CC immunodeficiencies, graft versus host disease, AIDS, bronchial asthma,  
 CC Crohn's disease or multiple sclerosis. They are also useful for treating  
 CC angiogenic-associated disorders including psoriasis, wound healing or  
 CC stroke. The present sequence encodes a NOVX protein.  
 XX  
 SQ Sequence 2363 BP; 456 A; 696 C; 808 G; 403 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2363; DB 12; Length 2363;  
 Best Local Similarity 100.0%; Pred. NO. 0; Mismatches 0; Gaps 0;  
 Matches 2363; Conservative 0; Indels 0; Gaps 0;  
 QY 1 TCGAGCCGCTTTTCAGGAGCCTTACCTGAGGGCCACAGGTGAGGAGCCTGAGCTTAC 60  
 DB 2363 TCGAGCCGCTTTTCAGGAGCCTTACCTGAGGGCCACAGGTGAGGAGCCTGAGCTTAC 2304  
 QY 61 AGGCCCAAGCCAGCGCTTGCCTCCAGGCGCGCCGCTGCTGGGGGCCACCATGCTCC 120  
 DB 2303 AGGCCCAAGCCAGCGCTTGCCTCCAGGCGCGCCGCTGCTGGGGGCCACCATGCTCC 2244

QY 121 TGCCAGAGCTTGAGAGCTGACCCGAGCCCGGACCTACCTGAGAGCTCCGCCACCTGC 180  
 DB 2243 TGCCAGAGCTTGAGAGCTGACCCGAGCCCGGACCTACCTGAGAGCTCCGCCACCTGC 2184  
 QY 181 TGAGCCCAAGGATGAGGAGCAAGAGGCCCGGACCTGACAGTTCCAGCCCTGAGAGCAGAGG 240  
 DB 2183 TGAGCCCAAGGATGAGGAGCAAGAGGCCCGGACCTGACAGTTCCAGCCCTGAGAGCAGAGG 2124  
 QY 241 TTCCTCATATCCCCCAGCCAGCTTAATAGCCCACTCTTAATAGAGGGGTTCTGGGGAGC 300  
 DB 2123 TTCCTCATATCCCCCAGCCAGCTTAATAGCCCACTCTTAATAGAGGGGTTCTGGGGAGC 2064  
 QY 301 TGAAGAGGGGACATATGAGTCTGCCCAAGACCTAGTGTCTGTCTGCTTCTTCTT 360  
 DB 2063 TGAAGAGGGGACATATGAGTCTGCCCAAGACCTAGTGTCTGTCTGCTTCTTCTT 2004  
 QY 361 CAGACTGAGCGTTTGAAGCCCAAGTCTTTCCTGCCAGAGCCAGAGTTCCAGCCCTCAG 420  
 DB 2003 CAGACTGAGCGTTTGAAGCCCAAGTCTTTCCTGCCAGAGCCAGAGTTCCAGCCCTCAG 1944  
 QY 421 GCCCTCTCTCTCATATCTAGGAGTCTGGGCCCCCAATTTCTCTTCTTCCAAAGATT 480  
 DB 1943 GCCCTCTCTCTCATATCTAGGAGTCTGGGCCCCCAATTTCTCTTCTTCCAAAGATT 1884  
 QY 481 TGATTTGAGGTCTCAGCTGTCTCTCTCTCAACCGGAGTCTCAGTCCCTGCTCTCAC 540  
 DB 1883 TGATTTGAGGTCTCAGCTGTCTCTCTCTCTCAACCGGAGTCTCAGTCCCTGCTCTCAC 1824  
 QY 541 CAGGCTGAGCATGGGGGTTCCCATCTCTGCAAAATTCAGGCGTCCCGCTGCTGTCA 600  
 DB 1823 CAGGCTGAGCATGGGGGTTCCCATCTCTGCAAAATTCAGGCGTCCCGCTGCTGTCA 1764  
 QY 601 GACACTGACCCCATCTCTTAAACCGGCAATCTGCTCCGTATACAGCGCTCTCTGG 660  
 DB 1763 GACACTGACCCCATCTCTTAAACCGGCAATCTGCTCCGTATACAGCGCTCTCTGG 1704  
 QY 661 CCAAGGCGCATGCTCTTACAGCTCTGCTGATGAGACGCTGAGGCGCCAGGAGCT 720  
 DB 1703 CCAAGGCGCATGCTCTTACAGCTCTGCTGATGAGACGCTGAGGCGCCAGGAGCT 1644  
 QY 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCTCCCGTCAATCTCTCAAGTCCACCC 780  
 DB 1643 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCTCCCGTCAATCTCTCAAGTCCACCC 1584  
 QY 781 TGGCCCAAGAGTCAAGCAGGAGATCTTAAACAAGAGCGTACATGGGCGAGAGAG 840  
 DB 1583 TGGCCCAAGAGTCAAGCAGGAGATCTTAAACAAGAGCGTACATGGGCGAGAGAG 1524  
 QY 841 GGTGGCCGAGTGTGCATCTGCTCCAGACCCAGGTGGCAGCTCTCACTGGCGGAGCC 900  
 DB 1523 GGTGGCCGAGTGTGCATCTGCTCCAGACCCAGGTGGCAGCTCTCACTGGCGGAGCC 1464  
 QY 901 CTGTACTTGTGACAGCATGCGGGGCGCATCTGTGGCCATGTGTGCTTTCTTCAG 960  
 DB 1463 CTGTACTTGTGACAGCATGCGGGGCGCATCTGTGGCCATGTGTGCTTTCTTCAG 1404  
 QY 961 AGTGAACAGAGAGCGCTGTATCCAGTGCAGAGTCAAGCTTGGCGCTCATGAGTC 1020  
 DB 1403 AGTGAACAGAGAGCGCTGTATCCAGTGCAGAGTCAAGCTTGGCGCTCATGAGTC 1344  
 QY 1021 TTGACAAAGACGAGAGGAGCGTGGCGGCTGTGTCTCTCGGCTTCAACGCGAGGGTA 1080  
 DB 1343 TTGACAAAGACGAGAGGAGCGTGGCGGCTGTGTCTCTCGGCTTCAACGCGAGGGTA 1284  
 QY 1081 GCGGACTCAGCTGCGAGAGAGATGGGCTTCTCAAGGGCACTGACCCGAGCTGAGAC 1140  
 DB 1283 GCGGACTCAGCTGCGAGAGAGATGGGCTTCTCAAGGGCACTGACCCGAGCTGAGAC 1224  
 QY 1141 GTGGAACCGCGGCGGCGCAATGGCACTGGGCTTCTTCTGTGTGACAGAGGGAGGCTG 1200  
 DB 1223 GTGGAACCGCGGCGGCGCAATGGCACTGGGCTTCTTCTGTGTGACAGAGGGAGGCTG 1164  
 QY 1201 CCCACACCCAGAGGCTGTGAGAGTCACTCCGTGTGTATGGCCCAAGAGCCGTTTC 1260

```
Db 1163 CCCACACCCAGAGGCTCTGAGGTCATCTCCGTTGTGATGTGCCCCAGAGGCCGTTTC 1104
Qy 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGAGAAAGCTGCCGTGTGACCGCATGTGGGA 1320
Db 1103 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGAGAAAGCTGCCGTGTGACCGCATGTGGGA 1044
Qy 1321 GGGCCGGGACACCAAGTTGGGCGGCTGGGCGGTGGCAAGTCAGCTTTCGCTATGATGAGACA 1380
Db 1043 GGGCCGGGACACCAAGTTGGGCGGCTGGGCGGTGGCAAGTCAGCTTTCGCTATGATGAGACA 984
Qy 1381 CACCTCTGTGGGGGATCCCTGCTCTCCGGGACCTGGGCTGTGACAGCCGCCAATGCTTC 1440
Db 983 CACCTCTGTGGGGGATCCCTGCTCTCCGGGACCTGGGCTGTGACAGCCGCCAATGCTTC 924
Qy 1441 CCGAGCGGAAACCGGGCTCTGTCCCGATGGCGAGTTTGGCCGGTCCGTTGAGCC 1500
Db 923 CCGAGCGGAAACCGGGCTCTGTCCCGATGGCGAGTTTGGCCGGTCCGTTGAGCC 864
Qy 1501 TCTCCCGCAGGCTCTGACCTGGGGGTGCAAGCTGTGTCTACACGGGGGCTATCTTCC 1560
Db 863 TCTCCCGCAGGCTCTGACCTGGGGGTGCAAGCTGTGTCTACACGGGGGCTATCTTCC 804
Qy 1561 TTTGGGAGCCCAACAGGAGAGAAAGCAACGATATTGGCCGTGGTCACTCTCCAGT 1620
Db 803 TTTGGGAGCCCAACAGGAGAGAAAGCAACGATATTGGCCGTGGTCACTCTCCAGT 744
Qy 1621 CCCCTGCCCCCTACAGAAATACAGCTGCTGTGTGCTCCAGCTGCGGCGAGGCCCTG 1680
Db 743 CCCCTGCCCCCTACAGAAATACAGCTGCTGTGTGCTCCAGCTGCGGCGAGGCCCTG 684
Qy 1681 GTGATGGCAAGATCTGTACCTGTGACGGGCTGGGGCAACAGCACTATGCGCAACAG 1740
Db 683 GTGATGGCAAGATCTGTACCTGTGACGGGCTGGGGCAACAGCACTATGCGCAACAG 624
Qy 1741 GCGGGGGATCTCCAGAGGCTCGAGTCCCAATTAATCAGCANATGTGTGAATGGGCT 1800
Db 623 GCGGGGGATCTCCAGAGGCTCGAGTCCCAATTAATCAGCANATGTGTGAATGGGCT 564
Qy 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTCTGTGTGCTGCTACCCGAGGGTGC 1860
Db 563 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTCTGTGTGCTGCTACCCGAGGGTGC 504
Qy 1861 ATTATGCTGTGCGCAGGCGGACAGCGGGTGTCCCTTGTGTGTGAGAGACATCTTCGG 1920
Db 503 ATTATGCTGTGCGCAGGCGGACAGCGGGTGTCCCTTGTGTGTGAGAGACATCTTCGG 444
Qy 1921 ACGCAGGTTGGCGGCTGTGTGCAATTGTGAGTTGGGCACTGTGCTGTGCCCTGGCCAG 1980
Db 443 ACGCAGGTTGGCGGCTGTGTGCAATTGTGAGTTGGGCACTGTGCTGTGCCCTGGCCAG 384
Qy 1981 AAGCAGGCGCTCTACACCAAAATGATGACTTCCGGGAGTGGATCTTTCAGGGCATAAG 2040
Db 383 AAGCAGGCGCTCTACACCAAAATGATGACTTCCGGGAGTGGATCTTTCAGGGCATAAG 324
Qy 2041 ACTCACTCCGAAGCCAGCGGATGTGACCCAGCTCTGACCGGGTGTCTTCTGCTGGCA 2100
Db 323 ACTCACTCCGAAGCCAGCGGATGTGACCCAGCTCTGACCGGGTGTCTTCTGCTGGCA 264
Qy 2101 GCGTCAGGGGCGAGGATCCCGGTGTGGATCCAGCGTGGGCGAGAGATGGGACGT 2160
Db 263 GCGTCAGGGGCGAGGATCCCGGTGTGGATCCAGCGTGGGCGAGAGATGGGACGT 204
Qy 2161 TTTTCTTTTGGGCGGCTCCAGAGTCCAGAGACACCTCTCCAGAGGCTCTCTCTTC 2220
Db 203 TTTTCTTTTGGGCGGCTCCAGAGTCCAGAGACACCTCTCCAGAGGCTCTCTCTTC 144
Qy 2221 CACAGTGGCGGCGCACTAGCGCCGAGACCAACCACTCACTCTCTGACCCCATGT 2280
Db 143 CACAGTGGCGGCGCACTAGCGCCGAGACCAACCACTCACTCTCTGACCCCATGT 84
Qy 2281 AAATATTGTCTGCTGTGGAGTCCGCTAGAGTCCCTGATGATGGATGCTCTT 2340
```

```
Db 83 AAATATTGTCTGCTGTGGAGTCCGCTGTAGTGGCCCTGATGATGGATGCTCTT 24
Qy 2341 AAATATTAAAGATGTTTGAAT 2363
Db 23 AAATATTAAAGATGTTTGAAT 1
RESULT 8
ACC46451
ID ACC46451 standard; cDNA; 2311 BP.
XX
AC ACC46451;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human dlthp protein modification/maintenance protein-encoding cDNA.
XX
KW Human; dlthp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW protein modification; protein maintenance; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200297031-A2.
XX
PD 05-DEC-2002.
XX
PE 27-MAR-2002; 2002WO-US010056.
XX
PF 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291282P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuson O, Yap PE, Amshy SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RX, Urashka ME;
XX
DR WPI: 2003-129518/12.
XX
PT P-PSDB: ABR41513.
XX
XX
PS Claim 2; SEQ ID NO 372; 591bp; English.
XX
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dlthp (ACC46080-ACC46749) and to their encoded
XX proteins (DLTHP; ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dlthp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dlthp nucleic acid sequence; the
XX recombinant production of DLTHP proteins; antibodies specific for DLTHP
XX proteins; microarrays comprising dlthp nucleic acid sequences; methods of
XX detecting dlthp nucleotide and protein sequences; methods of screening
XX for compounds that specifically bind a DLTHP protein; and methods of
XX assessing the toxicity of test compounds using a dlthp hybridisation
XX probe. Dithp nucleic acid sequences and DLTHP proteins may be used in the
```









|    |      |   |      |
|----|------|---|------|
| OY | 736  | CCCCCAAGGCTCGGCTCTCCCGGCTCATCTCTCTACAGAGTCCACCTCGGCCAGAGGTCA      | 795  |
| Db | 157  | CTTCAGAGGCTCGCCCCCACTCGCTGGAACCCA-GGATCCCACTTCGCGCCAGAGGTCA       | 215  |
| OY | 796  | GCACAGGATCACTTAAACAAGAGCATGACATGAGGCGAAGAGAGGTGGCCGACCTG          | 855  |
| Db | 216  | GCCAGGGAAATCATTAACAAGAGGACATGACATGGCGCGAAGAGAGAGGTGGCCGACCTG      | 275  |
| OY | 856  | CCATGCTGCTCCAGACCCCAAGGTGGACGTCCTACCTGGGAGACCTGTACTTTTGACA        | 915  |
| Db | 276  | CCATGCTGCTCCAGACCCCAAGGTGGACGTCCTACCTGGGAGACCTGTACTTTTGACA        | 335  |
| OY | 916  | GCCATCGGGGCGGCATCTTGAGGCACTTGATGCTGTTCTCTCAGAGTGAACAGAGCCG        | 975  |
| Db | 336  | GCCATCGGGGCGGCATCTTGAGGCACTTGATGCTGTTCTCTCAGAGTGAACAGAGCCG        | 395  |
| OY | 976  | CTGTATCCAGTSCAGTSCAGCTCTGCGGACGCTCGGCTCATGAGTCTTTTGAACAAGCGAA     | 1035 |
| Db | 396  | CTGTATCCAGTSCAGTSCAGCTCTGCGGACGCTCGGCTCATGAGTCTTTTGAACAAGCGAA     | 455  |
| OY | 1036 | GAGGACGTGGCGGCTGCTGCTGCTCTCGGCGCTCAACGCGCAGGGTAGCGGACTCAGCTGC     | 1095 |
| Db | 456  | GAGGACGTGGCGGCTGCTGCTGCTCTCGGCGCTCAACGCGCAGGGTAGCGGACTCAGCTGC     | 515  |
| OY | 1096 | GAGGACGTGGCTTCTCAAGGCACTGACCCATCTCGACTGGAAGTGCAGAACGCGGGGCT       | 1155 |
| Db | 516  | GAGGACGTGGCTTCTCAAGGCACTGACCCATCTCGACTGGAAGTGCAGAACGCGGGGCT       | 575  |
| OY | 1156 | GCCATATGGCACTCGGGCTTCTTCTGTGTGACACGAGGGAGGCTGCCCCACCCAGAG         | 1215 |
| Db | 576  | GCCATATGGCACTCGGGCTTCTTCTGTGTGACACGAGGGAGGCTGCCCCACCCAGAG         | 635  |
| OY | 1216 | CTGCTGAGAGTCACTCTCCGTGTGTATTTGCCCCAAGGCGGTTTCTTGAGCGGCATCTGC      | 1275 |
| Db | 636  | CTGCTGAGAGTCACTCTCCGTGTGTATTTGCCCCAAGGCGGTTTCTTGAGCGGCATCTGC      | 695  |
| OY | 1276 | CAGAAGTGTGGCCGCGAGGAAGCTGCGCGTGAACGCACTGTGTGAGGACCGGACACCAAGC     | 1335 |
| Db | 696  | CAGAAGTGTGGCCGCGAGGAAGCTGCGCGTGAACGCACTGTGTGAGGACCGGACACCAAGC     | 755  |
| OY | 1336 | TTTGGGCGCGGTGGCCCGTGGCAAGTCAGGCTTGCGCTATGATGAGACACACTCTGTGTGGGA   | 1395 |
| Db | 756  | TTTGGGCGCGGTGGCCCGTGGCAAGTCAGGCTTGCGCTATGATGAGACACACTCTGTGTGGGA   | 815  |
| OY | 1396 | TCCCTGCTCTCCGAGGACCTGGGTGTGTACAGCCGCGCCCATGCTTCCGGAAGCGGAAACCGG   | 1455 |
| Db | 816  | TCCCTGCTCTCCGAGGACCTGGGTGTGTACAGCCGCGCCCATGCTTCCGGAAGCGGAAACCGG   | 875  |
| OY | 1456 | GTCCTGTGCCGATGTGGCGAGTGTTCGCCGAGTCCGTCAGGCTCTCTCCACGATCTG         | 1515 |
| Db | 876  | GTCCTGTGCCGATGTGGCGAGTGTTCGCCGAGTCCGTCAGGCTCTCTCCACGATCTG         | 935  |
| OY | 1516 | CAGCTGTGGGGGTGGCAGGCTGTGTGTCTACACAGGGGGCTATCTTCCCTTTGGGACCCCAAC   | 1575 |
| Db | 936  | CAGCTGTGGGGGTGGCAGGCTGTGTGTCTACACAGGGGGCTATCTTCCCTTTGGGACCCCAAC   | 995  |
| OY | 1576 | AGCGAGGAGAACAGCAACGATATTGCCCTGTGGCCACTCTCAGTCCCTGACCCCTTACCA      | 1635 |
| Db | 996  | AGCGAGGAGAACAGCAACGATATTGCCCTGTGGCCACTCTCAGTCCCTTACCAAGTCCCTTACCA | 1055 |
| OY | 1636 | GAATATCATCCAGCTGTGTGCCCTCCACAGCTGCGGCGCAGGCGCTGTGTGATGTGCAAGATC   | 1695 |
| Db | 1056 | GAATATCATCCAGCTGTGTGCCCTCCACAGCTGCGGCGCAGGCGCTGTGTGATGTGCAAGATC   | 1115 |
| OY | 1696 | TGTATCCGTGAAGGGGTGTGGGGCAACACGCAAGTACTATGGCCAAACAAGCCGGGGTACTCCAG | 1755 |
| Db | 1116 | TGTATCCGTGAAGGGGTGTGGGGCAACACGCAAGTACTATGGCCAAACAAGCCGGGGTACTCCAG | 1175 |
| OY | 1756 | GAGGCTCAGTCTCCCATATACGACATATGATGTCTGCAATGTGGCGGCTTCTTAATGGAAC     | 1815 |
| Db | 1176 | GAGGCTCAGTCTCCCATATATACGACATATGATGTCTGCAATGTGGCGGCTTCTTAATGGAAC   | 1235 |
| OY | 1816 | CAGATCAAGCCCAAGATGTTCTGTGCTGAGTCAACCCGAGGGGTGCAATTGATGCTGCGAG     | 1875 |

|           |   |  |      |
|-----------|---|--|------|
| Db        | 1236  | CAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCGAGGGGTGGCATTTGATGCTTCGACG  | 1295 |
| Qy        | 1876  | GGCCGACAGCGGTGATCCCTTTGTGTGTGAGGACAGACATCTCTGGACGCCAGTTGGCCGG  | 1935 |
| Db        | 1236  | GGCCGACAGCGGTGATCCCTTTGTGTGTGAGGACAGACATCTCTGGACGCCAGTTGGCCGG  | 1355 |
| Qy        | 1936  | CTGTGTGGCATTTGATGTGGGGGACATGGCTGTGCCCCGTGGCCAGAAAGCCAGGCGTCTAC | 1995 |
| Db        | 1356  | CTGTGTGGCATTTGATGTGGGGGACATGGCTGTGCCCCGTGGCCAGAAAGCCAGGCGTCTAC | 1415 |
| Qy        | 1996  | ACCAAAAGTCAATGACTTTCGGGAAGTGATCTTCCAGGCGATTAAGACTCACTCCGAAGCC  | 2055 |
| Db        | 1416  | ACCAAAAGTCAATGACTTTCGGGAAGTGATCTTCCAGGCGATTAAGACTCACTCCGAAGCC  | 1475 |
| Qy        | 2056  | AGCCGCAATGTGACCCCAAGCTCTGACCGGTGGCTTCTGGCTGGCCAGGCTCCAGGGCCCGA | 2115 |
| Db        | 1476  | AGCCGCAATGTGACCCCAAGCTCTGACCGGTGGCTTCTGGCTGGCCAGGCTCCAGGGCCCGA | 1535 |
| Qy        | 2116  | GGTGTATCCCGGTGTGGGATCCACGCTGGGCCGAGATGGGACGTTTTTCTTCTTGGGCC    | 2175 |
| Db        | 1536  | GGTGTATCCCGGTGTGGGATCCACGCTGGGCCGAGATGGGACGTTTTTCTTCTTGGGCC    | 1595 |
| Qy        | 2176  | CGGTTCACAGGTCCAAAGACACCTCTCCATGAGGTCTCTTCCACAGTGGCGGCCCC       | 2235 |
| Db        | 1596  | CGGTTCACAGGTCCAAAGACACCTCTCCATGAGGTCTCTTCCACAGTGGCGGCCCC       | 1655 |
| Qy        | 2236  | ACTAGAGCCCGAGACACCCCAACTCAACCTCCGACGCCCCAGTAAATATTTGCTGCT      | 2295 |
| Db        | 1656  | ACTAGAGCCCGAGACACCCCAACTCAACCTCCGACGCCCCAGTAAATATTTGCTGCT      | 1715 |
| Qy        | 2296  | GTCGTGGAGCTCCTCTTACGATGCCCCCTGATGATGGATGCTTTTAAATTAATAAGATGG   | 2355 |
| Db        | 1716  | GTCGTGGAGCTCCTCTTACGATGCCCCCTGATGATGGATGCTTTTAAATTAATAAGATGG   | 1775 |
| Qy        | 2356  | TTTTTGATT 2363   |      |
| Db        | 1776  | TTTTGATT 1783  |      |
| RESULT 11 |   |  |      |
| AA643103  |   |  |      |
| ID        | AA643103  | standard; cDNA; 1783 BP.                                       |      |
| XX        | AC  | AA643103;  |      |
| XX        | DT  | 19-NOV-2002 (first entry)                                      |      |
| XX        | DE  | cDNA encoding Hepsin.  |      |
| XX        | KM  | Human; cancer; hepsin; hepsin protease; gene; ss.              |      |
| OS        | Homo sapiens.   |  |      |
| XX        | PN  | WO200162271-A1.  |      |
| XX        | PD  | 30-AUG-2001.   |      |
| XX        | PF  | 20-FEB-2001; 2001WO-US005703.                                  |      |
| XX        | PR  | 22-FEB-2000; 2000US-00510738.                                  |      |
| PA        | (UVAR-)   | UNIV ARKANSAS.   |      |
| XX        | PI  | O'Brien TJ;  |      |
| XX        | DR  | WPI; 2001-582004/65.   |      |
| PT        | New oligonucleotide complementary to hepsin encoding sequence, useful for |  |      |
| PT        | treating cancer and screening for compounds that inhibit hepsin.          |  |      |
| PS        | Claim 35; 138-139; 126p; English.   |  |      |
| XX        |   |  |      |

The invention relates to an oligonucleotide (i) having a complementary sequence to hepsin. The oligonucleotide is useful for the detection of cancer, treatment of cancer and screening for compounds that inhibit hepsin activity. Hepsin protease, mRNA and immunospecific anti-hepsin antibodies are useful for the diagnosis of cancer in an individual. CC AAS43076-AAS43103 represent coding sequences of cancer in an individual. CC the methods of the invention

SO Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;

Query Match 67.4%; Score 1593.6; DB 5; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 7.2e-313;

Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCAGGCGCTGCTCCCTCCATCTCTCAAGGTCACACCTGCGCCAGAGGTCA 795  
DB 157 CTTGAGGCTCCGCCCCCAGCTGTCGACCCCA-GGGTCCACCTGCGCCAGAGGTCA 215  
QY 796 GCCAGGATCATTAACAAGAGGAGTGCATGCGCGAGAGAGAGGGTGGCCGAGCTGTG 855  
DB 216 GCCAGGATCATTAACAAGAGGAGTGCATGCGCGAGAGAGAGGGTGGCCGAGCTGTG 275  
QY 856 CCATGCTGCTCAGACCCCAAGTGGAGACTCTCACTGCGGGAGCCCTGCTACTTCTGACA 915  
DB 276 CCATGCTGCTCAGACCCCAAGTGGAGACTCTCACTGCGGGAGCCCTGCTACTTCTGACA 335  
QY 916 GCCATGCGGGGCGGATCTGGGCGCATTTGGGCTGTTCTCCAGAGGTGACAGAGACCG 975  
DB 336 GCCATGCGGGGCGGATCTGGGCGCATTTGGGCTGTTCTCCAGAGGTGACAGAGACCG 395  
QY 976 CTGTACCCAGTGCAGTCTGCGGACGCTCGGCTCATGTGCTTTTGAACAAGCGAA 1035  
DB 396 CTGTACCCAGTGCAGTCTGCGGACGCTCGGCTCATGTGCTTTTGAACAAGCGAA 455  
QY 1036 GGGACGTGGCGGCTGCTGTGCTCTCGGCTCCAAACGCGAGGTAGCGGACTCAGCTGC 1095  
DB 456 GGGACGTGGCGGCTGCTGTGCTCTCGGCTCCAAACGCGAGGTAGCGGACTCAGCTGC 515  
QY 1096 GAGAGATGGGCTTCTCAAGGCGCATGCCATCCAGCTGAGCGAGCGGAGCGGCGG 1155  
DB 516 GAGAGATGGGCTTCTCAAGGCGCATGCCATCCAGCTGAGCGAGCGGAGCGGCGG 575  
QY 1156 GCCAATGGCAAGTGGGCTTCTTGTGTGAGACGAGGAGGAGCTGCCCAACCCAGAGG 1215  
DB 576 GCCAATGGCAAGTGGGCTTCTTGTGTGAGACGAGGAGGAGCTGCCCAACCCAGAGG 635  
QY 1216 CTGCTGGAAGTCACTCCGTGTGTGATTGGCCCAAGAGCGGTTCTTGGCCGCTGTC 1275  
DB 636 CTGCTGGAAGTCACTCCGTGTGTGATTGGCCCAAGAGCGGTTCTTGGCCGCTGTC 695  
QY 1276 CAAGACTGTGGCGGCAAGAGCTGCCGCTGGAACGCGCATTCGAGGAGCGGGAACACAGC 1335  
DB 696 CAAGACTGTGGCGGCAAGAGCTGCCGCTGGAACGCGCATTCGAGGAGCGGGAACACAGC 755  
QY 1336 TTGGGCGGCTGGCGGCTGGAAGCTGCGCTTGCATGATGAGACACCTCTGTGGGGA 1395  
DB 756 TTGGGCGGCTGGCGGCTGGAAGCTGCGCTTGCATGATGAGACACCTCTGTGGGGA 815  
QY 1396 TCCCTGCTCTCCGAGGACTGGGTCTGACAGCGCGCCACTGCTTCCGAGCGGAACCGG 1455  
DB 816 TCCCTGCTCTCCGAGGACTGGGTCTGACAGCGCGCCACTGCTTCCGAGCGGAACCGG 875  
QY 1456 GTCCGTGCCGAGTGGGAGTGTGGCGGTCGAGGCGGTCAGAGGCTCTCCCAAGGCTG 1515  
DB 876 GTCCGTGCCGAGTGGGAGTGTGGCGGTCGAGGCGGTCAGAGGCTCTCCCAAGGCTG 935  
QY 1516 CAGCTGGGAGTGCAGGCTGTGTGCTACACGAGGAGCTATCTTCCCTTGGGACCCCAAC 1575  
DB 936 CAGCTGGGAGTGCAGGCTGTGTGCTACACGAGGAGCTATCTTCCCTTGGGACCCCAAC 995  
QY 1576 AGCGAGAGAAACAGCAAGATATGCTGCTGCTCACTCTTCAGTCCCTGCGCCCTGACA 1635  
DB 996 AGCGAGAGAAACAGCAAGATATGCTGCTGCTCACTCTTCAGTCCCTGCGCCCTGACA 1055

QY 1636 GAATACATTCAGGCTGTGTGCTCTCCAGCTGCGGCGGAGCCCTGCTGTGATGCGAAGATC 1695  
DB 1056 GAATACATTCAGGCTGTGTGCTCTCCAGCTGCGGCGGAGCCCTGCTGTGATGCGAAGATC 1115  
QY 1696 TGTACCGTGAAGGCTGGGCAACACGACAGTACTATGCGCAAGCGCGGAGTACTCAG 1755  
DB 1116 TGTACCGTGAAGGCTGGGCAACACGACAGTACTATGCGCAAGCGCGGAGTACTCAG 1175  
QY 1756 GAGGCTGAGTCCCATATACAGCATGATGTCTGCATGCGCTGACTTCTATGGAAC 1815  
DB 1176 GAGGCTGAGTCCCATATACAGCATGATGTCTGCATGCGCTGACTTCTATGGAAC 1235  
QY 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGAGTGGCATTTGATGCGCAG 1875  
DB 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGAGTGGCATTTGATGCGCAG 1295  
QY 1876 GAGCAGACGCGTGGTCCCTTGTGTGTGAGAGACAGATCTCTCGAAGCGGCTGGCGG 1935  
DB 1296 GAGCAGACGCGTGGTCCCTTGTGTGTGAGAGACAGATCTCTCGAAGCGGCTGGCGG 1355  
QY 1936 CTGTGTGGCATTTGTGATTTGGGCACTGGCTGTGCTTGGCCAGAGCCAGGCGTTAC 1995  
DB 1356 CTGTGTGGCATTTGTGATTTGGGCACTGGCTGTGCTTGGCCAGAGCGGCTGTAC 1415  
QY 1996 ACCAAGTCAAGTACTTCCGAGAGTGGATCTTCCAGGCGCATTAAGACTCACTCGAAGCC 2055  
DB 1416 ACCAAGTCAAGTACTTCCGAGAGTGGATCTTCCAGGCGCATTAAGACTCACTCGAAGCC 1475  
QY 2056 AGCGGCAATGATGACCCAGCTCTGACCGGTGGCTTCTGCTGCGCAGGCTTCCAGGCGCGA 2115  
DB 1476 AGCGGCAATGATGACCCAGCTCTGACCGGTGGCTTCTGCTGCGCAGGCTTCCAGGCGCGA 1535  
QY 2116 GGTGATCCCGTGTGTGATTCACAGCTGGGCGGAGATGGGAAGTTTCTTCTTGGGCC 2175  
DB 1536 GGTGATCCCGTGTGTGATTCACAGCTGGGCGGAGATGGGAAGTTTCTTCTTGGGCC 1595  
QY 2176 CCGTCAACAGTGCACAGACACCTCCCTCAGGAGTCTCTTCCACAGTGGCGGCGCC 2235  
DB 1596 CCGTCAACAGTGCACAGACACCTCCCTCAGGAGTCTCTTCCACAGTGGCGGCGCC 1655  
QY 2236 ACTCAGCCCGGAGACCAACCACTCACTCTGACCCCGATGTAATATTTGCTGCT 2295  
DB 1656 ACTCAGCCCGGAGACCAACCACTCACTCTGACCCCGATGTAATATTTGCTGCT 1715  
QY 2296 GTCTGGAGCTCTGTCTAGGTGCCCTGATGATGGATGCTCTTAAATTAATGAATGG 2355  
DB 1716 GTCTGGAGCTCTGTCTAGGTGCCCTGATGATGGATGCTCTTAAATTAATGAATGG 1775  
QY 2356 TTTTGATT 2363  
DB 1776 TTTTGATT 1783

RESULT 12  
AB079532  
ID AB079532 standard; cDNA; 1783 BP.

XX AB079532;  
XX  
XX 25-NOV-2002 (first entry)

DE Human hepsin polypeptide encoding cDNA.

XX Hepsin; cancer; tumour; cytosolic; gene therapy; human; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 246..1499

FT /tag= a

XX /product= "hepsin"

PN WO200264839-A2.  
XX  
XX 22-AUG-2002.  
XX  
XX 12-FEB-2002; 2002WO-US004018.  
XX  
XX 14-FEB-2001; 2001US-0268361P.  
XX  
XX (TULIA-) TULIARIC INC.  
XX  
XX Mu D, Powers S;  
XX WPI; 2002-643496/69.  
XX P-P8DB; ABB81197.  
XX  
XX Diagnosing a cancer in a mammal, useful for preventing and treating  
XX cancer, e.g. ovarian, prostate, breast or lung cancer, comprises  
XX detecting and measuring the hepsin gene copy number and comparing the  
XX data obtained to a control data.  
XX  
XX Claim 25; Page 77; 77pp; English.  
XX  
XX The invention relates to diagnosing cancer in a mammal by detecting the  
XX hepsin gene. The method involves detecting and measuring the hepsin gene  
XX copy number in a biological subject from a region of the mammal that is  
XX suspected to be precancerous or cancerous to generate data for a test  
XX gene copy, and comparing the test gene copy number to a data for a  
XX control gene copy number, where an amplification of the gene in the  
XX biological subject relative to the control indicates the presence of a  
XX precancerous lesion or cancer in the animal. The method is useful for  
XX diagnosing, preventing and treating cancer, e.g. ovarian, prostate,  
XX breast or lung cancer, by the detection of the hepsin gene. The hepsin  
XX genes and their expressed protein products are useful as targets for  
XX therapy and for identifying compounds useful in the diagnosis, prevention  
XX and treatment of tumours and cancers. The present sequence represents the  
XX human hepsin polypeptide encoding cDNA (GenBank Accession No. M18930)  
XX  
SQ Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;  
Query Match 67.4%; Score 1593.6; DB 6; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 7.2e-313;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
QY 736 CCCCAGGCTCTGCTCCCTCCATCTCTCAAGGTGCCACCTGGCCCAAGAGTCA 795  
DB 157 CCTGAGAGCTCCGCCCACTGCTGACCCCA-GGGTCCCACTGGCCCAAGAGTCA 215  
QY 796 GCCAGGATCATTAACAAGAGGAGTGAATGGGCGAAGAGAGGGGAGGAGTGTG 855  
DB 216 GCCAGGATCATTAACAAGAGGAGTGAATGGGCGAAGAGAGGGGAGGAGTGTG 275  
QY 856 CCATGCTCTCAAGACCAAGAGTGAATGGGCGAAGAGGAGGAGTGTGCTCACTTCA 915  
DB 276 CCATGCTCTCAAGACCAAGAGTGAATGGGCGAAGAGGAGGAGTGTGCTCACTTCA 335  
QY 916 GCCATGGGAGGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGAATGGAGCCG 975  
DB 336 GCCATGGGAGGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGAATGGAGCCG 395  
QY 976 CTGATACCAAGTGAATGGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGAATGGAG 1035  
DB 396 CTGATACCAAGTGAATGGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGAATGGAG 455  
QY 1036 GGAAGTGGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGTGC 1095  
DB 456 GGAAGTGGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGTGC 515  
QY 1096 GAGAGATGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1155  
DB 516 GAGAGATGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 575  
QY 1156 GCCAATGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1215

DB 576 GCCAATGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 635  
QY 1216 CTGCTGAGAGTGAATGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1275  
DB 636 CTGCTGAGAGTGAATGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 695  
QY 1276 CAAGAGTGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1335  
DB 696 CAAGAGTGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 755  
QY 1336 TTGGGCGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1395  
DB 756 TTGGGCGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 815  
QY 1396 TCCCTGCTCTCCGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 1455  
DB 816 TCCCTGCTCTCCGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 875  
QY 1456 GTCCGTGCTCTCCGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 1515  
DB 876 GTCCGTGCTCTCCGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 935  
QY 1516 CAGCTGGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1575  
DB 936 CAGCTGGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 995  
QY 1576 AGCAGAGAGAGAGAGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 1635  
DB 996 AGCAGAGAGAGAGAGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCA 1055  
QY 1636 GAATACATCAAGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1695  
DB 1056 GAATACATCAAGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1115  
QY 1696 TGTACCTGAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1755  
DB 1116 TGTACCTGAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1175  
QY 1756 GAGGCTGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1815  
DB 1176 GAGGCTGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1235  
QY 1816 CAGATCAAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1875  
DB 1236 CAGATCAAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1295  
QY 1876 GAGCAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1935  
DB 1296 GAGCAGAGGAGTGTGCTCTCAAGAGTGAATGGGCGAAGAGTGTGCTCTCAAGAGTGC 1355  
QY 1936 CTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1995  
DB 1356 CTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1415  
QY 1996 ACCAAGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2055  
DB 1416 ACCAAGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1475  
QY 2056 AGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2115  
DB 1476 AGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1535  
QY 2116 GGTGATCCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2175  
DB 1536 GGTGATCCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1595  
QY 2176 CGGTTCACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2235  
DB 1596 CGGTTCACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1655  
QY 2236 ACTCAGCCCGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2295  
DB 1656 ACTCAGCCCGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1715

```

Oy      2226  GTCTGGGACTCTGCTCTAGTGGCCCCCTCATGATGGGAGTCCCTTTAAATTAATAAGATGG 2355
Db      1716  GTCTGGGACTCTGCTCTAGTGGCCCCCTCATGATGGGAGTCCCTTTAAATTAATAAGATGG 1775
Oy      2356  TTTTGATT 2363
Db      1776  TTTTGATT 1783

RESULT 13
ID      ADAL17045
XX      ADAL17045 standard; cDNA; 1783 BP.
AC      ADAL17045;
XX
DT      06-NOV-2003 (first entry)
XX
DE      cDNA encoding human hepsin.
XX
KW      Malignant hyperplasia detection method; hepsin protease; protease;
KW      cancer; ovarian cancer; tumour cell; ovary; lung; prostate; colon;
XX      cystostatic; human; gene; ss.
XX
OS      Homo sapiens.
XX
PN      US2003027181-A1.
PD      06-FEB-2003.
XX
PF      20-MAR-2002; 2002US-00102283.
XX
PR      22-FEB-2000; 2000US-00510738.
PR      21-MAY-2001; 2001US-00861966.
PR      30-JUL-2001; 2001US-00919048.
XX
PA      (OBRI/) O'BRIEN T J.
PA      (CANN/) CANNON M J.
PA      (SANT/) SANTIN A.
XX
PI      O'Brien TJ, Cannon MJ, Santin A;
DR      WPI; 2003-531460/50.
XX
PT      Detecting malignant hyperplasia in a biological sample e.g. blood, urine,
PT      saliva, tears, by isolating mRNA from the sample, and detecting hepsin
PT      mRNA in the sample for the presence or absence of the mRNA in the sample.
XX
PS      Disclosure; Page 48-49; 74pp; English.
XX
CC      The present invention relates to a method for detecting malignant
CC      hyperplasia in a biological sample. The method comprises isolating mRNA
CC      from the sample, and detecting hepsin mRNA in the sample, where the
CC      presence or absence of hepsin mRNA in the sample is indicative of the
CC      presence or absence of malignant hyperplasia. Also disclosed are PCR
CC      primers which can be used to detect the presence of specific proteases in
CC      a tissue sample. The detected proteases, specifically hepsin protease,
CC      are over-expressed in certain cancers such as ovarian cancer. The method
CC      of the invention is useful for detecting malignant hyperplasia in a
CC      biological sample such as blood, urine, saliva, tears, interstitial
CC      fluid, ascites fluid, tumour tissue biopsy and circulating tumour cells.
CC      Also disclosed are methods for inhibiting the expression of endogenous
CC      hepsin in a cell, and for inhibiting the hepsin protein in a cell. The
CC      method is useful for targeted therapy to an individual, where the
CC      individual suffers from a cancer selected from ovary, lung, prostate, and
CC      colon. The present sequence encodes human hepsin.
XX
SQ      Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;

Query Match      67.4%; Score 1593.6; DB 9; Length 1783;
Best Local Similarity 99.1%; Pred. No. 7.2e-313;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1,

```

|    |      |   |      |
|----|------|---|------|
| QY | 736  | CCCCAAGCCCTGCTCTCCCGTCATCTCTCACAGGTCCCAACCTTGGCCCCAAGAGTCA      | 795  |
| Db | 157  | CCCTGAGAGCTCCGCCCCACCTGTGTGAGCCCCA - GGCTCCCACTCTGGCCAGAGGTCA   | 215  |
| QY | 796  | GCCAGGGAAATCATTAACAAGAGGCAATGACATGACCGCAAGAGAGGGTGGCCGAGACTGTG  | 855  |
| Db | 216  | GCCAGGGAAATCATTTAAACAAGGCAATGACATGACCGCAAGAGAGGGTGGCCGAGACTGTG  | 275  |
| QY | 856  | CCATGCTGCTCCAGACCCAAGTGGCAGCTCTCACTGGGGAGCCCTGCTACTTCTTGACA     | 915  |
| QY | 276  | CCATGCTGCTCCAGACCCAAGTGGCAGCTCTCACTGGGGAGCCCTGCTACTTCTTGACA     | 335  |
| QY | 916  | GCCATCGGGCGGATCTCTGGGCCATTGTGGCTGTTCTCTCAGAGATGACAGAGACCG       | 975  |
| Db | 336  | GCCATCGGGCGGATCTCTGGGCCATTGTGGCTGTTCTCTCAGAGATGACAGAGACCG       | 395  |
| QY | 976  | CTGATACCAGTGCAGGTCAAGTCTGTCGGGAGCGCTCCGCTCAATGATCTTTGACAAGCGAA  | 1035 |
| Db | 396  | CTGATACCAGTGCAGGTCAAGTCTGTCGGGAGCGCTCCGCTCAATGATCTTTGACAAGCGAA  | 455  |
| QY | 1036 | GGAGCGTGGCGGCTGCTGTGCTCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCAAGCTGC   | 1095 |
| Db | 456  | GGAGCGTGGCGGCTGCTGTGCTCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCAAGCTGC   | 515  |
| QY | 1096 | GAGAGATTGGGCTTCTTCAGAGGCACTGACCACTCCGAGCTGGACGTGTGCAACAGCGGGC   | 1155 |
| Db | 516  | GAGAGATTGGGCTTCTTCAGAGGCACTGACCACTCCGAGCTGGACGTGTGCAACAGCGGGC   | 575  |
| QY | 1156 | GCCAATGGCAAGTCGGGCTCTCTGTGTGTGAAGAGAGGGAGGGTCCCAACCCAGAG        | 1215 |
| Db | 576  | GCCAATGGCAAGTCGGGCTCTCTGTGTGTGAAGAGAGGGAGGGTCCCAACCCAGAG        | 635  |
| QY | 1216 | CTGCTGAGGCTCATCTCCGTGTGTGATTTGCCAGAGGCCGTTCCTTGGCCGCATCTGC      | 1275 |
| Db | 636  | CTGCTGAGGCTCATCTCCGTGTGTGATTTGCCAGAGGCCGTTCCTTGGCCGCATCTGC      | 695  |
| QY | 1276 | CAAGACTGTGGCCGCGAGAAAGCTGCCCTGTGACCGCATCTGTGGAGGCCGAGACACAGC    | 1335 |
| Db | 696  | CAAGACTGTGGCCGCGAGAAAGCTGCCCTGTGACCGCATCTGTGGAGGCCGAGACACAGC    | 755  |
| QY | 1336 | TTGGGCGGATGGGCGGTGGCAAGTCAGGCTTGGCATGATGAGAGACAACCTGTGGGGAG     | 1395 |
| Db | 756  | TTGGGCGGATGGGCGGTGGCAAGTCAGGCTTGGCATGATGAGAGACAACCTGTGGGGAG     | 815  |
| QY | 1396 | TCCCTGCTCTCCGGGGAGCTGGGTGTCTGACAGCGGCCCATCTGTTCCCGAGCCGAAACCGG  | 1455 |
| Db | 816  | TCCCTGCTCTCCGGGGAGCTGGGTGTCTGACAGCGGCCCATCTGTTCCCGAGCCGAAACCGG  | 875  |
| QY | 1456 | GTCCTGTCCCGATGGCGAGTGTGTCGGGTGCGGTGGCCAGGCTCTCCCAAGGTCTG        | 1515 |
| Db | 876  | GTCCTGTCCCGATGGCGAGTGTGTCGGGTGCGGTGGCCAGGCTCTCCCAAGGTCTG        | 935  |
| QY | 1516 | CAGCTGGGGGTGGCAGGCTGTGTGCTACAGGGGGGCTATCTTCCCTTGGGGAGCCCAAC     | 1575 |
| Db | 936  | CAGCTGGGGGTGGCAGGCTGTGTGCTACAGGGGGGCTATCTTCCCTTGGGGAGCCCAAC     | 995  |
| QY | 1576 | AGCAGAGAGAAACAGCAACGATATTGCGCTGTCAACTCTTCAGTCCCTGCCCCCTACA      | 1635 |
| Db | 996  | AGCAGAGAGAAACAGCAACGATATTGCGCTGTCAACTCTTCAGTCCCTGCCCCCTACA      | 1055 |
| QY | 1636 | GAATATCATCCAGCCTGTGTGCTCTCCAGCTGCCGGCCAGGCTCTGTGTGATGGCAAGTC    | 1695 |
| Db | 1056 | GAATATCATCCAGCCTGTGTGCTCTCCAGCTGCCGGCCAGGCTCTGTGTGATGGCAAGTC    | 1115 |
| QY | 1696 | TGTATCCGTGAACGGGCTTGGGGCAACAGCAATTAATAAGGCCAACAAGCCGGGGTAACTCAG | 1755 |
| Db | 1116 | TGTATCCGTGAACGGGCTTGGGGCAACAGCAATTAATAAGGCCAACAAGCCGGGGTAACTCAG | 1175 |
| QY | 1756 | GAGGCTGAGTCCCATTAATCAGAAATGATGTGCAATGGCGCTGACTCTTAATGGAAC       | 1815 |
| Db | 1176 | GAGGCTGAGTCCCATTAATCAGAAATGATGTGCAATGGCGCTGACTCTTAATGGAAC       | 1235 |
| QY | 1816 | CAGATCAAGCCCAAGATGTTCTGTGTGCGCTAACCCGAGGGGTGCAATTGATGCTGCGAG    | 1875 |

Db 1236 CAGATCAAGCCCAAGATTTCTGTGCTGCTACCCCGAGGGTGGCATTTGATGCTCCAG 1235  
 Qy 1876 GGCACACAGCGGTGTCCTTTGTGTGTGAGACAGCATCTCTCCAGCCGACGTTGGCG 1935  
 Db 1296 GGCACACAGCGGTGTCCTTTGTGTGTGAGACAGCATCTCTCCAGCCGACGTTGGCG 1355  
 Qy 1336 CTGTGTGCGATTTGAGTTGGGGCACTGGCTGTGCTCCCGGCGAAGACCGGGGTCTAC 1935  
 Db 1356 CTGTGTGCGATTTGAGTTGGGGCACTGGCTGTGCTCCCGGCGAAGACCGGGGTCTAC 1415  
 Qy 1996 ACCAAGCTAGTGACTTCCGGAGTGAATCTTCCAGCGCCATAAAGACTCACTCCGAAGC 2055  
 Db 1416 ACCAAGCTAGTGACTTCCGGAGTGAATCTTCCAGCGCCATAAAGACTCACTCCGAAGC 1475  
 Qy 2056 AGCGGCATGTGACCCAGCTCTGACCCGGTGTCTCTGCTGCGACGCTCCAGGCGCCGA 2115  
 Db 1476 AGCGGCATGTGACCCAGCTCTGACCCGGTGTCTCTGCTGCGACGCTCCAGGCGCCGA 1535  
 Qy 2116 GGTATATCCCGTGTGTGGATTCACGCTGGGCGGAGATGGACGTTTTTCTTTGGGCC 2175  
 Db 1536 GGTATATCCCGTGTGTGGATTCACGCTGGGCGGAGATGGACGTTTTTCTTTGGGCC 1595  
 Qy 2176 CGGTCCACAGGTCCAGACACACCTCTCCAGGGTCTCTCTTCCACAGTGGCGGCC 2235  
 Db 1596 CGGTCCACAGGTCCAGACACACCTCTCCAGGGTCTCTCTTCCACAGTGGCGGCC 1655  
 Qy 2236 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACCCCGCATGTAATATTTGTTGCT 2295  
 Db 1656 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACCCCGCATGTAATATTTGTTGCT 1715  
 Qy 2296 GTCTGGGACTCTGTCTTAGTGTGCTGATGTAGTGGATGCTTTAATAATAAAGATG 2355  
 Db 1716 GTCTGGGACTCTGTCTTAGTGTGCTGATGTAGTGGATGCTTTAATAATAAAGATG 1775  
 Qy 2356 TTTTGATT 2363  
 Db 1776 TTTTGATT 1783

RESULT 14  
 ADB75353  
 ID ADB75353 standard; cDNA; 1783 BP.  
 XX  
 AC ADB75353;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Prostate cancer marker cDNA.  
 XX  
 KW Prostate; cancer; cytostatic; gene therapy; marker; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO2003009814-A2.  
 XX  
 PD 06-FEB-2003.  
 XX  
 PF 25-JUL-2002; 2002WO-US023913.  
 XX  
 PR 25-JUL-2001; 2001US-0307982P.  
 PR 22-AUG-2001; 2001US-0314356P.  
 PR 25-SEP-2001; 2001US-0325020P.  
 PR 12-DEC-2001; 2001US-0341746P.  
 PR 05-MAR-2002; 2002US-0362158P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Schlegel R, Monahan JE, Endegge WO, Gannavarapu M, Gorbacheva B,  
 Hoeroh S, Kamatkar S, Womsey AM, Glatt K, Zhao X, Anderson D,  
 WPI; 2003-248033/24.  
 XX

PT New nucleic acid molecule, useful for diagnosing or treating prostate  
 cancer.  
 XX  
 PS Disclosure; SEQ ID NO 177; 999P; English.  
 XX  
 CC The invention relates to newly discovered cancer markers associated with  
 CC the cancerous state of prostate cells. Also disclosed is a method of  
 CC assessing whether a patient is afflicted with prostate cancer. The method  
 CC of the invention involves assessing whether a patient is afflicted with  
 CC prostate cancer by comparing the level of expression of a marker in a  
 CC patient sample and the normal level of expression of the marker in a  
 CC control non-prostate cancer sample, where a significant increase in the  
 CC level of expression of the marker in the patient sample and the normal  
 CC level indicates that the patient is afflicted with prostate cancer.  
 CC Nucleic acids of the invention are useful for diagnosing or treating  
 CC prostate cancer, and may be useful in gene therapy. Sequences given in  
 CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;  
 XX  
 Query Match 67.4%; Score 1593.6; DB 10; Length 1783;  
 Best Local Similarity 99.1%; Pred. No. 7.2e-313;  
 Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 Qy 736 CCCCCAGGCCCTGCTCCGCTCCATCTCTCAACAGTCCACCTGCGCCAGAGGTCA 795  
 Db 157 CTTGAGAGCTCCGCCCCCACTGTGAGCCCA-GGGTCCACCTGCGCCAGAGGTCA 215  
 Qy 796 GCCAGGGAATATTAAAGAGGCAATGACATGCGCGAAGAGGTGCGGACTGTG 855  
 Db 216 GCCAGGGAATATTAAAGAGGCAATGACATGCGCGAAGAGGTGCGGACTGTG 275  
 Qy 856 CCATGCTCTCCAGACCCAGAGTGGAGCTCTCACTCGGAGGACCTGTCTACTTGACA 915  
 Db 276 CCATGCTCTCCAGACCCAGAGTGGAGCTCTCACTCGGAGGACCTGTCTACTTGACA 335  
 Qy 916 GCCATCGGGGCGGATCTGGGCGCATTTGGCTGTTCTCTCAGAGTGAACAGAGCCG 975  
 Db 336 GCCATCGGGGCGGATCTGGGCGCATTTGGCTGTTCTCTCAGAGTGAACAGAGCCG 395  
 Qy 976 CTGTACCAAGTGAAGTCACTCTGCGGAGGCTGTGCTCATGTCTTTGACAGAGCGAA 1035  
 Db 396 CTGTACCAAGTGAAGTCACTCTGCGGAGGCTGTGCTCATGTCTTTGACAGAGCGAA 455  
 Qy 1036 GGGACGTGGGCGGCTGTGCTCTCTGCGGCTCCAGCGCCAGGAGGCGGACTCAGCTGC 1095  
 Db 456 GGGACGTGGGCGGCTGTGCTCTCTGCGGCTCCAGCGCCAGGAGGCGGACTCAGCTGC 515  
 Qy 1096 GAGAGATGGGCTTCTCTCAGAGGCACTGACCCATCTCGAGTGAAGTGTGAGACGCGGCG 1155  
 Db 516 GAGAGATGGGCTTCTCTCAGAGGCACTGACCCATCTCGAGTGTGAGACGCGGCGCG 575  
 Qy 1156 GCCAATGGCACTGTGGGCTTCTCTGTGTGAGAGAGGGAGGCTGCCCAACCCAGAGG 1215  
 Db 576 GCCAATGGCACTGTGGGCTTCTCTGTGTGAGAGAGGGAGGCTGCCCAACCCAGAGG 635  
 Qy 1216 CTGCTGAGGTCATCTCCGTTGTGATTTGCGCCAGAGGCGGTTCTTGGGCGGATCTGC 1275  
 Db 636 CTGCTGAGGTCATCTCCGTTGTGATTTGCGCCAGAGGCGGTTCTTGGGCGGATCTGC 695  
 Qy 1276 CAAGACTGTGGCGCGAGAGAGTGCCTCGGTGAGCCGATCTGTGGAGGCGCGGACACGAGC 1335  
 Db 696 CAAGACTGTGGCGCGAGAGAGTGCCTCGGTGAGCCGATCTGTGGAGGCGCGGACACGAGC 755  
 Qy 1336 TTGGGCGGTTGGCGCGGAGTCAAGCTTTCGCTTAGTATGAGACACCTCTGTGGGGGA 1395  
 Db 756 TTGGGCGGTTGGCGCGGAGTCAAGCTTTCGCTTAGTATGAGACACCTCTGTGGGGGA 815  
 Qy 1396 TCCCTGCTCCGCGGAGCTGGGTGTGACAGCGCGCCACTGTCTTCCGAGAGCGGAACCGG 1455

```
Db      816  TCCCTGCTCTCCGGGACTGAGTGTGCTGACAGCCGCCCACTCTTCCGGAGCGGAACCGG  875
Qy      1456  GTCTGTCCCGATGGCGAGTGTGGCCGCTGCGGTGCGCCGAGGCTCTCCCCAAGGTCTG  1515
Db      876  GTCTGTCCCGATGGCGAGTGTGGCCGCTGCGGTGCGCCGAGGCTCTCCCCAAGGTCTG  935
Qy      1516  CAGCTGGGGGTGACAGGCTGTGTCTACCAACGGGGGTATCTTCCCTTCCGGAGCCCAAC  1575
Db      936  CAGCTGGGGGTGACAGGCTGTGTCTACCAACGGGGGTATCTTCCCTTCCGGAGCCCAAC  995
Qy      1576  AGCGAGGAGAAACAGCAAGATATTGCCCTGTGTCACCTCTCAGTCCCTGCGCCCTACA  1635
Db      996  AGCGAGGAGAAACAGCAAGATATTGCCCTGTGTCACCTCTCAGTCCCTGCGCCCTACA  1055
Qy      1636  GAATACATCCAGCTGTGTGCTCTCCAGCTGCGCGGCGCCCTGTGTGATGGCAATATC  1695
Db      1056  GAATACATCCAGCTGTGTGCTCTCCAGCTGCGCGGCGCCCTGTGTGATGGCAATATC  1115
Qy      1696  TGTACCGTGAACGGCTGGGGGACACGACGACTATGGCAACAGGCGGGGTACTCCAG  1755
Db      1116  TGTACCGTGAACGGCTGGGGGACACGACGACTATGGCAACAGGCGGGGTACTCCAG  1175
Qy      1756  GAGGCTCGAGTCCCATTAATCAGCAATGATGTGCAATGGCGCTGACTTCTATGAAAC  1815
Db      1176  GAGGCTCGAGTCCCATTAATCAGCAATGATGTGCAATGGCGCTGACTTCTATGAAAC  1235
Qy      1816  CAGATCAAGCCCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1875
Db      1236  CAGATCAAGCCCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1295
Qy      1876  GGCACAGACGGGTGTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1935
Db      1296  GGCACAGACGGGTGTCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1355
Qy      1936  CTGTGTGTCATTGTGAGTTGGGAGCACTGGCTGTGTGCTGTGCTGTGCTGTGCTGTGCT  1995
Db      1356  CTGTGTGTCATTGTGAGTTGGGAGCACTGGCTGTGTGCTGTGCTGTGCTGTGCTGTGCT  1415
Qy      1996  ACCAAATGTCAGTGACTTCCGGGAGTGTGATCTTCCAGGCAATAAAGACTCACTCCGAA  2055
Db      1416  ACCAAATGTCAGTGACTTCCGGGAGTGTGATCTTCCAGGCAATAAAGACTCACTCCGAA  1475
Qy      2056  AGCGGCATGTGACCCAGCTGTGACCGGTGTGCTTCTGCTGTGCGAGCTCCAGGGCCGA  2115
Db      1476  AGCGGCATGTGACCCAGCTGTGACCGGTGTGCTTCTGCTGTGCGAGCTCCAGGGCCGA  1535
Qy      2116  GGTGATCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  2175
Db      1536  GGTGATCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1595
Qy      2176  CGGTCCACAGTCCAAAGACACCTCTCCAGGCTCTCTTCCAGAGTGGGGGCC  2235
Db      1596  CGGTCCACAGTCCAAAGACACCTCTCCAGGCTCTCTTCCAGAGTGGGGGCC  1655
Qy      2236  ACTCAGCCCCGAGACCAACCACTCAACCTCTCTGACCCCGCATTAATATTGTCTGCT  2295
Db      1656  ACTCAGCCCCGAGACCAACCACTCAACCTCTCTGACCCCGCATTAATATTGTCTGCT  1715
Qy      2296  GTCTGGGACTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  2355
Db      1716  GTCTGGGACTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1775
Qy      2356  TTTTGATT 2363
Db      1776  TTTTGATT 1783
```

```
RESULT 15
ADD18429
ID      ADD18429 standard; DNA; 1783 BP.
XX
XX      ADD18429;
XX
```

```
DT      15-JAN-2004 (first entry)
XX
DE      Human prostate cancer diagnosis related DNA sequence SegID1.
XX
KW      prostate tissue; cancer diagnostic; cancer marker; prostate cancer; PCA;
KW      male cancer-related death; serum biomarker; tissue biomarker; cytostatic;
KW      gene therapy; prostate biopsy tissue; AMACR;
KW      alpha-methylacyl-coenzyme A racemase; diagnosing cancer; cell growth;
KW      human; de.
XX
OS      Homo sapiens.
XX
PN      WO2003012067-A2.
XX
PD      13-FEB-2003.
XX
PF      02-AUG-2002; 2002MO-US024567.
XX
PR      02-AUG-2001; 2001US-0309581P.
PR      15-NOV-2001; 2001US-0334468P.
PR      01-AUG-2002; 2002US-00210120.
XX
PA      (UNMI ) UNIV MICHIGAN.
XX
PI      Rubin MA, Chinnaiyan AM, Sreekumar A;
XX
DR      WPI; 2003-278396/27.
XX
XX      Characterizing prostate tissue comprises providing a prostate tissue
PT      sample from a subject and detecting the presence or absence of expression
PT      of hepsin, pim-1 or EZH2.
XX
PS      Disclosure; SEQ ID NO 1; 297bp; English.
XX
CC      This invention relates to a novel method of characterizing prostate
CC      tissue in a subject and to compositions and methods for cancer
CC      diagnostics, including cancer markers, in particular prostate cancer.
CC      Prostate cancer (PCA) is a leading cause of male cancer-related death.
CC      Additional serum and tissue biomarkers would aid diagnosis. The invention
CC      may provide means of producing compounds with a cytostatic activity or
CC      allow the development of gene therapy. The methods of the invention
CC      useful for characterizing prostate tissue in a subject, screening
CC      compound, characterizing inconclusive prostate biopsy tissue in a
CC      subject, detecting AMACR (alpha-methylacyl-coenzyme A racemase)
CC      expression in a bodily fluid, characterizing tissue in a subject,
CC      diagnosing cancer in a subject and inhibiting the growth of cells. The
CC      present sequence is a DNA sequence which is preferably utilised in the
CC      method of the invention.
XX
SQ      Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;
XX
Query Match      67.4%; Score 1593.6; DB 10; Length 1783;
Best Local Similarly 99.1%; Pred. No. 7.2e-313;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy      736  CCCCAGAGCCCTGCTCCCGCTCATCTCTCAAGGTCCACCTGGCCAGAGGTCA  795
Db      157  CTTGAGGCTTCGCCCCCACTGTGACCCCA-GGGTCCACCTTGCCAGAGGTCA  215
Qy      796  GCCAGGGAATCATTAACAAGAGCAATGACATGGCGCAGAAGAGGGTGCCTGTG  855
Db      216  GCCAGGGAATCATTAACAAGAGCAATGACATGGCGCAGAAGAGGGTGCCTGTG  275
Qy      856  CCATGCTGTCCAGACCCCAAGGTGTGACGTCTCACTGCGGGGACCCCTGTACTTCTACA  915
Db      276  CCATGCTGTCCAGACCCCAAGGTGTGACGTCTCACTGCGGGGACCCCTGTACTTCTACA  335
Qy      916  GCCATCGGGGCGGATCTGCGGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  975
Db      336  GCCATCGGGGCGGATCTGCGGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  395
Qy      976  CTGTACCCAGTGAAGTCACTCTGCGGACGCTCGGCTCATGTGTCTTTTGAACAAGGAA  1035
Db      396  CTGTACCCAGTGAAGTCACTCTGCGGACGCTCGGCTCATGTGTGTCTTTTGAACAAGGAA  1035
```





***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 02:14:07, Search time 11629 Seconds  
(without alignments)  
7734.618 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363  
Sequence: 1 tcggagccgcgttcacagga.....taataaagatggtttgattc 2363

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gest1.\*  
9: gb\_gest2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match Length | ID     | Description        |
|------------|--------|--------------------|--------|--------------------|
| 1          | 1238   | 52.4               | 1238 9 | AY408324 Homo sapi |
| 2          | 1060.8 | 44.9               | 1225 9 | AY408325 Pan trogl |
| 3          | 1057.6 | 44.8               | 1828 3 | CR597177 full-leng |
| 4          | 939.2  | 39.7               | 1814 3 | AK002694 Mus muscu |
| 5          | 932.4  | 39.5               | 1238 9 | AY408326           |
| 6          | 905.4  | 38.3               | 1212 3 | CR592189 full-leng |
| 7          | 831    | 35.2               | 846 1  | AL558100           |
| 8          | 814    | 34.4               | 916 4  | BG754441           |
| 9          | 796.8  | 33.7               | 940 4  | BG754531           |
| 10         | 794    | 33.6               | 965 5  | BK406458           |
| 11         | 788.2  | 33.4               | 1012 7 | CK231686           |
| 12         | 766.4  | 32.4               | 881 4  | BT682846           |
| 13         | 753    | 31.9               | 961 5  | BK422288           |
| 14         | 747    | 31.6               | 874 5  | BK427780           |
| 15         | 726.8  | 29.9               | 774 4  | BT752774           |
| 16         | 705.6  | 29.8               | 887 4  | BT760912           |
| 17         | 704.4  | 29.9               | 717 4  | BT759565           |
| 18         | 702.6  | 29.7               | 910 7  | CN156776           |
| 19         | 680.6  | 28.8               | 829 7  | CR765441           |
| 20         | 671.8  | 28.4               | 837 5  | BK325135           |
| 21         | 671.6  | 28.4               | 786 5  | BK406457           |
| 22         | 668    | 28.3               | 688 5  | BM994658           |
| 23         | 668    | 28.3               | 931 7  | CO775189           |
| 24         | 665.2  | 28.2               | 701 7  | CV029569           |

|   |    |       |      |      |   |          |          |           |
|---|----|-------|------|------|---|----------|----------|-----------|
| c | 25 | 660   | 27.9 | 690  | 5 | BU688738 | BU688738 | UI-CF-EC1 |
|   | 26 | 659.8 | 27.9 | 763  | 4 | BT764518 | BT764518 | 603050810 |
|   | 27 | 652   | 27.6 | 911  | 5 | B0645325 | B0645325 | AGENCOURT |
|   | 28 | 640.6 | 27.1 | 830  | 4 | BT517769 | BT517769 | 603042013 |
|   | 29 | 635.8 | 26.9 | 649  | 1 | AV656299 | AV656299 | AV656299  |
|   | 30 | 622.8 | 26.4 | 786  | 4 | BT765615 | BT765615 | 603043887 |
|   | 31 | 618.6 | 26.2 | 776  | 4 | BT836887 | BT836887 | 603089619 |
|   | 32 | 616.6 | 26.1 | 698  | 4 | CD614168 | CD614168 | 560596943 |
|   | 33 | 616.4 | 26.1 | 984  | 4 | BT762697 | BT762697 | 603048453 |
|   | 34 | 612.8 | 25.9 | 812  | 7 | CN154671 | CN154671 | 941901 MA |
|   | 35 | 603.8 | 25.6 | 1142 | 4 | BMS48649 | BMS48649 | AGENCOURT |
|   | 36 | 603.4 | 25.5 | 628  | 6 | CD614167 | CD614167 | 560596944 |
|   | 37 | 599.8 | 25.4 | 609  | 7 | CN272681 | CN272681 | 170006000 |
|   | 38 | 596   | 25.2 | 738  | 5 | BR444180 | BR444180 | BP444180  |
|   | 39 | 593.6 | 25.1 | 930  | 1 | AL528314 | AL528314 | U19508.Y  |
|   | 40 | 590.8 | 25.0 | 681  | 2 | AM008459 | AM008459 | WVS5601.x |
|   | 41 | 570   | 24.1 | 879  | 4 | BT764734 | BT764734 | 603043363 |
|   | 42 | 563.6 | 23.9 | 651  | 1 | AL561243 | AL561243 | CQ27706.x |
|   | 43 | 560.8 | 23.7 | 1116 | 1 | AL557891 | AL557891 | AL557891  |
|   | 44 | 558.2 | 23.6 | 564  | 1 | AL699708 | AL699708 | DKEP2686B |
|   | 45 | 552.6 | 23.4 | 777  | 4 | BT766513 | BT766513 | 603052314 |

ALIGNMENTS

|                            |  |             |     |        |                 |
|----------------------------|--|-------------|-----|--------|-----------------|
| RESULT 1                   | AY408324   | 1238 bp     | DNA | linear | GSS 15-DEC-2003 |
| LOCUS                      | AY408324   | 1238 bp     | DNA | linear | GSS 15-DEC-2003 |
| DEFINITION                 | Homo sapiens HPN gene, VIRTUAL TRANSCRIPT, partial sequence.   |             |     |        |                 |
| ACCESSION                  | AY408324   |             |     |        |                 |
| VERSION                    | AY408324.1   | GI:39764295 |     |        |                 |
| KEYWORDS                   | GSS.   |             |     |        |                 |
| SOURCE                     | Homo sapiens (human)   |             |     |        |                 |
| ORGANISM                   | Homo sapiens   |             |     |        |                 |
| REFERENCE                  | Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. |             |     |        |                 |
| AUTHORS                    | Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. |             |     |        |                 |
| TITLE                      | Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  |             |     |        |                 |
| COMMENT                    | This sequence was made by sequencing genomic exons and ordering them based on alignment.   |             |     |        |                 |
| FEATURES                   | location/Qualifiers  |             |     |        |                 |
| source                     | 1..1238  |             |     |        |                 |
| gene                       | /organism="Homo sapiens"   |             |     |        |                 |
|                            | /mol_type="genomic DNA"  |             |     |        |                 |
|                            | /db_xref="taxon:9606"  |             |     |        |                 |
|                            | <1..>1238  |             |     |        |                 |
|                            | /gene="HPN"  |             |     |        |                 |
|                            | /locus_tag="HCM3186"   |             |     |        |                 |
| ORIGIN                     |  |             |     |        |                 |
| Query Match                | 52.4%; Score 1238; DB 9; Length 1238;  |             |     |        |                 |
| Best Local Similarity      | 100.0%; Pred. No. 1.9e-277;  |             |     |        |                 |
| Matches 1238; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;   |             |     |        |                 |
| 842                        | GTGGCGGACGTCGATGCTGCTCGACACCGAAGTGGAGCTCTGACGCGGAGCC   |             |     |        |                 |
| 901                        | 901  |             |     |        |                 |

Db 1 GTGGCCGGAGCTGTGCATGTGCTTCAGAACCCAAAGTGGCAGCTCTCACTCGGGGAGACC 60  
 Qy 902 TGTACTTCTTGAAGCCATTCGGGGGGGAGATCTCTGGGCGCATTTGTGGCTGTCTCTCAGGA 961  
 Db 61 TGTACTTCTTGAAGCCATTCGGGGGGGAGATCTCTGGGCGCATTTGTGGCTGTCTCTCAGGA 120  
 Qy 962 GTGACAGAGAGCCGCTGTACCCAGTGCAGTCACTGTCCGGAAGCTCGGCTCAATGCTCT 1021  
 Db 121 GTGACAGAGAGCCGCTGTACCCAGTGCAGTCACTGTCCGGAAGCTCGGCTCAATGCTCT 180  
 Qy 1022 TTGAACAAG 1081  
 Db 181 TTGAACAAG 240  
 Qy 1082 CCGGACTAGCTGTGAG 1141  
 Db 241 CCGGACTAGCTGTGAG 300  
 Qy 1142 TGGCAACGGGGGGGCGCAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201  
 Db 301 TGGCAACGGGGGGGCGCAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 Qy 1202 CCCAACCCAG 1261  
 Db 361 CCCAACCCAG 420  
 Qy 1262 TGGCGCGCATCTGTCAAG 1321  
 Db 421 TGGCGCGCATCTGTCAAG 480  
 Qy 1322 GCGGAG 1381  
 Db 481 GCGGAG 540  
 Qy 1382 ACTCTGTGGGGAGATCCCTGTCTTCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441  
 Db 541 ACTCTGTGGGGAGATCCCTGTCTTCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 Qy 1442 CCGAG 1501  
 Db 601 CCGAG 660  
 Qy 1502 CTCCCAACGAGCTGTGAG 1561  
 Db 661 CTCCCAACGAGCTGTGAG 720  
 Qy 1562 TTGGGGAG 1621  
 Db 721 TTGGGGAG 780  
 Qy 1622 CCGTGGCGCGCTCAAG 1681  
 Db 781 CCGTGGCGCGCTCAAG 840  
 Qy 1682 TGGATGGAGAGATCTGTACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741  
 Db 841 TGGATGGAGAGATCTGTACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 Qy 1742 CCGGGGGTACTCCAG 1801  
 Db 901 CCGGGGGTACTCCAG 960  
 Qy 1802 ACTCTGTAG 1861  
 Db 961 ACTCTGTAG 1020  
 Qy 1862 TTGATGCTGTGAG 1921  
 Db 1021 TTGATGCTGTGAG 1080  
 Qy 1922 CGGCAAGTGGAG 1981  
 Db 1081 CGGCAAGTGGAG 1140

Qy 1982 AGCCAGGCGTCTACACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATAAGA 2041  
 Db 1141 AGCCAGGCGTCTACACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATAAGA 1200  
 Qy 2042 CTCATCCGGAAGCCAGCGGATGGTGAACCAAGCTCTGA 2079  
 Db 1201 CTCATCCGGAAGCCAGCGGATGGTGAACCAAGCTCTGA 1238  
 RESULT 2  
 AY408325  
 LOCUS  
 DEFINITION Pan troglodytes HPN gene, VIRUTUAL TRANSCRIPT, partial sequence.  
 ACCESSION AY408325  
 VERSION AY408325.1 GI:39764296  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pan troglodytes (chimpanzee)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 REFERENCE  
 1 (bases 1 to 1225)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
 Todd,M.A., Tanenbaum,D.M., Civejello,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 Direct Submision  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 TITLE  
 Inferring nonneutral evolution from human-chimp-mouse orthologous  
 gene trios  
 JOURNAL  
 Science 302 (5652), 1960-1963 (2003)  
 PUBMED  
 14671302  
 REFERENCE  
 2 (bases 1 to 1225)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
 Todd,M.A., Tanenbaum,D.M., Civejello,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 COMMENT  
 This sequence was made by sequencing genomic exons and ordering  
 them based on alignment.  
 FEATURES  
 source  
 1..1225  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 <1..>1225  
 /gene="HPN"  
 /locus\_tag="HCM3186"  
 ORIGIN  
 Query Match 44.9%; Score 1060.8; DB 9; Length 1225;  
 Best Local Similarity 86.7%; Pred. No. 3,7e-236;  
 Matches 1062; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
 Qy 855 GCATGTGCTCCAG 914  
 Db 1 GCATGTGCTCCAG 60  
 Qy 915 AGCCATGGGGGGCGGCATCTGGGGCATTTGTGGCTTCTCTCAGAGTGAACAGAGAGCC 974  
 Db 61 AGCCATGGGGGGCGGCATCTGGGGCATTTGTGGCTTCTCTCAGAGTGAACAGAGAGCC 120  
 Qy 975 GCTGTACCCAGTGCAGTCACTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1034  
 Db 121 GCTGTACCCAGTGCAGTCACTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 Qy 1035 AGGAGAGTGGGGGTGTGCTCTCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1094  
 Db 181 AGGAGAGTGGGGGTGTGCTCTCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 Qy 1095 CGAGAGATGGGCTTCTCAGAGGAGCACTGAACCACTCCGAGCTGGACGTGGCAACGGGGGG 1154



Db 643 CACCTGTGTGGGGATCCCTGCTCTCCGGGGAGTGGGTTGTGACAGCCGCCCATGTGCTT 702

Qy 1441 CCGGAGCGGAACCGGGTCTCTCTCCGATGGGAGTGTGGCGGTCCGTGGCCCAAGCC 1500

Db 703 CCGGAGCGGAACCGGGTCTCTCTCCGATGGGAGTGTGGCGGTCCGTGGCCCAAGCC 762

Qy 1501 TCTCCCAAGCTGTGAGCTGGGGGTGAGGCTGTGCTCAACAGGGGGCTATCTTCC 1560

Db 763 TCTCCCAAGCTGTGAGCTGGGGGTGAGGCTGTGCTCAACAGGGGGCTATCTTCC 822

Qy 1561 TTTGGGAGCCCAACAGGAGAGAAACGACATATGGCTGTGCTCACTCTCCAGT 1620

Db 823 TTTGGGAGCCCAACAGGAGAGAAACGACATATGGCTGTGCTCACTCTCCAGT 882

Qy 1621 CCCCTGCCCTCAAGAAATATATCCAGCTGTGTGCTCCAGCTGCCGGCCAGGCTTG 1680

Db 883 CCCCTGCCCTCAAGAAATATATCCAGCTGTGTGCTCCAGCTGCCGGCCAGGCTTG 942

Qy 1681 GTGATGGCAAGATCTGTACCGTGCAGCGGGCTGGGGCAACGACAGTACTATGCGCAACAG 1740

Db 943 GTGATGGCAAGATCTGTACCGTGCAGCGGGCTGGGGCAACGACAGTACTATGCGCAACAG 1002

Qy 1741 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATGATGTCTGCAATGGCGCT 1800

Db 1003 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATGATGTCTGCAATGGCGCT 1062

Qy 1801 GACTTTCTATGGAACCAAGTCAAGTCAAGTCTTCTGTGCTGCTACCCCGAAGGTGGC 1860

Db 1063 GACTTTCTATGGAACCAAGTCAAGTCAAGTCTTCTGTGCTGCTACCCCGAAGGTGGC 1122

Qy 1861 ATTGATGCTTCCAGAGGCGGAGCGGAGTGTGCTTGTGTGTGAGAGACAGATCTCTCG 1920

Db 1123 ATTGATGCTTCCAGAGGCGGAGCGGAGTGTGCTTGTGTGTGAGAGACAGATCTCTCG 1182

Qy 1921 ACGCCAGCTTGGCGGCTGTGTGTGAGATTTGGAGTGGGCACTGGCTGTGCCCTGGCCAG 1980

Db 1183 ACGCCAGCTTGGCGGCTGTGTGTGAGATTTGGAGTGGGCACTGGCTGTGCCCTGGCCAG 1242

Qy 1981 AAGCCAGGCTTCTACCAAGTCAAGTCACTTCCGGAGTGGATCTTCCAGGCCATTA-- 2038

Db 1243 AAGCCAGGCTTCTACCAAGTCAAGTCACTTCCGGAGTGGATCTTCCAGGCCATTAAG 1302

Qy 2039 ----- 2038

Db 1303 GTGAAAGTTGGGTCCAGATGGGAGCCAGGGTGGGAGCTTTGGGTGTCTAATGGGGGAG 1362

Qy 2039 ----- 2038

Db 1363 GGAGGAGAGATTTGTTTAGGAAACCTACGCTCAGGCTAGAAAGGGCCCCCTTGGG 1422

Qy 2039 ----- 2038

Db 1423 AACAGATGACTTTGAAGGTTCTCTGGGAAAGGAAGCAGTGTGGAGCTGTGAAGCT 1482

Qy 2039 ----- AGACTCACTCCGAAGCAGGAGGAGT 2064

Db 1483 CTGAGACTCGGGAGCCCCAGCTGTCTTTTCCCAAGTCACTCCGAAGCAGGAGGAGT 1542

Qy 2065 GTGACCCAGCTGTGAACCGGTGGCTTCTGTGCTGCGAGCTTCAGAGGCCCCAGGTGATCC 2124

Db 1543 GTGACCCAGCTGTGAACCGGTGGCTTCTGTGCTGCGAGCTTCAGAGGCCCCAGGTGATCC 1602

Qy 2125 GGTGTGGGATTCACGCTGGGCGGAGATGGGAGCTTTTCTTCTGTGGGCCCCGTCCACA 2184

Db 1603 GGTGTGGGATTCACGCTGGGCGGAGATGGGAGCTTTTCTTCTGTGGGCCCCGTCCACA 1662

Qy 2185 GGTGCAAGAGACACCTCCCTCAGAGGTCTCTCTTCCAGTGGGAGGCCCCACTCAGGCC 2244

Db 1663 GGTGCAAGAGACACCTCCCTCAGAGGTCTCTCTTCCAGTGGGAGGCCCCACTCAGGCC 1722

Qy 2245 CGAGACCAACCAACCTCAACCTCTCTGACCCCAATGTAATATGTCTGTCTGTGGAC 2304

Db 1723 CGAGACCAACCAACCTCAACCTCTCTGACCCCAATGTAATATGTCTGTCTGTGGAC 1782

Qy 2305 TCCGTCTAGGTGGCCCTGATGATGGGATGCTCTTTAAATATATTA 2350

Db 1783 TCCGTCTAGGTGGCCCTGATGATGGGATGCTCTTTAAATATATTA 1828

RESULT 4  
AK002694  
LOCUS  
DEFINITION  
MUS musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:061030A17 product:hepsin, full insert sequence.  
ACCESSION  
AK002694  
VERSION  
AK002694.1 GI:12832864  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
PUBMED  
99279253  
REFERENCE  
AUTHORS  
1  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE  
20499374  
PUBMED  
11042159

3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kishimoto, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaue, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Maruyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
TITLE  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL  
Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE  
20530913  
PUBMED  
11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
Nature 409, 685-690 (2001)  
REFERENCE  
AUTHORS  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL  
Nature 420, 563-573 (2002)  
REFERENCE  
AUTHORS  
6 (bases 1 to 1814)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okio, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

TITLE  
Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),





Db 1591 GGCTCCACACATGAGGCTCCATCATGAAATGTTTCTGCTCAGATCCAGTCCAGGGGTCCAA 1650  
QY 2192 GGAACACCTCCCTCCAGGGTCTCTCTTCCACAGTGGGGGCCCACTAGCCCCGAGACC 2251  
Db 1651 GG--ATGCTGATCCAGGAGCTTCTCTTCCACAGTGGGCCGCCCACTCAATCCAGAGGCC 1708  
QY 2252 ACCCACTCAGCCCTCCAGAGCCCCCATGTAAATATTTGTTGCTGTGCTGAGACTCTGTG 2311  
Db 1709 ATTTGGCTCAGCCCTCC--ACCCATGTAAATATTTACTCTGTCTCTGAGGGGGGCTCT 1765  
QY 2312 TAGGTGCCCTGATGATGAGATGCTCTTTAAATAATAAAGATGTTTT 2359  
Db 1766 AGGAGAGCCCT--TGTGCAATGCTCTTTAAATAATAAAGTGTTTT 1811

RESULT 5  
AY408326 1238 bp DNA linear GSS 15-DEC-2003  
LOCUS Mus musculus HPN gene, VIRtual TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY408326  
VERSION AY408326.1 GI:39764297  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1238)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Todd,M.A., Tenenbaum,D.M., Ciavella,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Interfering nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1238)  
AUTHORS Todd,M.A., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
Clark,A.G., Tenenbaum,D.M., Ciavella,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
location/Qualifiers  
1..1238  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..>1238  
/gene="HPN"  
/locus\_tag="HGM3186"

ORIGIN  
Query Match 39.5%; Score 932.4; DB 9; Length 1238;  
Bee Local Similarity 84.6%; Pred. No. 3.2e-206;  
Matches 1047; Conservative 0; Mismatches 191; Indels 0; Gaps 0;  
QY 842 GTGGCCGAGCTGTGCATGCTGTCCAGACCAAGTGGAGCTCTCACTCGGGAGACC 901  
Db 1 GTGGCCGAGCTGTGCATGCTGTCCAGACCAAGTGGAGCTCTCACTGTGGGTACC 60  
QY 902 TGCTACTTGTGACAGCCATCGGGGCGGATCTGTGGCCATTGTGGCTGTTCTCCAGA 961  
Db 61 TGCTGTTCTGACAGGCAATGGGGCGCGTCTGTGGCCATTGTGACATCTCACTGACGA 120  
QY 962 GTGACGAGAGCCGCTGTACCAAGTGCAGTCACTGTGCGGAGCGCTGTGGCTCATGCT 1021  
Db 121 GTGACGAGAGCCACTGTACCAAGTGCAGTCACTGTGCGGAGCGCTGTGGCTGT 180  
QY 1022 TTGACAAGACGGAAGGAGCTGTGGCGGCTGTGTGCTCTCGGCTCCAAAGCCAGGGTAG 1081

Db 181 TTGACAAGACGGAAGGAGCTGTGGCGGCTGTGTGCTCTCGGCTCCAAAGCCAGGGTAG 240  
QY 1082 CCGAGCTGAGCTGAGAGAGATGGGCTTCTCAGGGCACTGACCACTCCGAGCTGAGAC 1141  
Db 241 CAGGCTCGGCTGTGAGAGATGGGCTTCTCAGGGCTGTGCACTCGGAGCTGATG 300  
QY 1142 TTGCAAGCGGGGGGCGCAATGGAGCACTGCGGGCTTCTCTGTGTGAGAGAGGGAGCTGC 1201  
Db 301 TGCCCACTGGGGGGCGCAAGGACATCGGGCTTCTTTGGTGTGAGAGGGCGGACTTC 360  
QY 1202 CCCAACCAGAGGCTGTGAGAGTCACTTCGCTGTGTGATTTGCCCAAGAGCCGTTTCT 1261  
Db 361 CTGTGCTCAGAGGTTCTGAGATGTCACTCTGTATGTGACTGTCTAAGAGGCGGATTC 420  
QY 1262 TGCGCCCATCTGCAAGACTGTGGCGCGAGAAAGCTGCCGTGTGACCCGATCTGGGAG 1321  
Db 421 TGACTGTCCACTGTCCAAAGACTGTGGCGCGAGAAAGCTGCCGTGTGACCCGATTC 480  
QY 1322 GCCGGGACACGAGCTTGGGGCGGAGCGGTGGCAAGTCAAGCTTCGCTATGATGAGACAC 1381  
Db 481 GCCAGAGACGAGCTGTGGAAAGTGGCGGTGGAGTCAAGCTTCGCTATGATGAGACCC 540  
QY 1382 ACTCTGTGGGGGATCCCTGCTCTCCGGGACTGGGTGTGACAGCCGCCACTGTTC 1441  
Db 541 ACCTGTGGGGGGTCCCTGCTGTGTGGGACTGGGGTGTGACTGTGACATTTGCTTTC 600  
QY 1442 CGAGCGGAAACCGGCTCTCTCCGATGTGGAGTGTTCGGGCTGTGGCCGCTGAGGCT 1501  
Db 601 CAGAGCGGAAACCGGCTCTCTCTCGGTGGAGATTTGTGTGTGTGTGTGTGTGTGTGTGT 660  
QY 1502 CTCCCAAGGCTGTGAGCTGTGGGGTGTGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1561  
Db 661 CACCCAGT 720  
QY 1562 TTGGGAGCCCCCAAGGAGAGAGAGAGAGAGAGAGATATTTGCCCTGTCCACTTC 1621  
Db 721 TTGGAGACCTTACTATGAGAGAGAGAGAGAGAGAGATATTTGCCCTGTCCACTTC 780  
QY 1622 CCTGTCCCTTCAAGATATCATTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1681  
Db 781 CCTGTCCCTTCAAGATATCATTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840  
QY 1682 TTGATGGCAAGATGTGACCGGTGTGAGCGGCTGTGGGCAACAGCACTATATGAGCCACAG 1741  
Db 841 TTGATGGCAAGGTCTGTACTGTGACCGGCTGTGGGTAACACAGTTCTATGGCCACAGG 900  
QY 1742 CCGGGTACTTCAGAGGCTCGAGTCCCATATACAGCAATGATGTCTGCAATGGCGTG 1801  
Db 901 CTATGGTGTCCAAAGAGCCCGGTTCCATCATATACAGCAAGATTGCAACAGCCCCG 960  
QY 1802 ACTTCTATGGAACCAATCAAGTCAAGCCCAAGATGTTGTGTGTGTGTGTGTGTGTGTGT 1861  
Db 961 ACTTCTAAGGAATCAATCAAGTCAAGCCCAAGATGTTGTGTGTGTGTGTGTGTGTGTGT 1020  
QY 1862 TTGATGCTGTGCAAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1921  
Db 1021 TTGATGCTGTGCAAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
QY 1922 CGCCACGTTGAGCGGCTGTGTGAGCAATGTGATTTGGGCACTGTGCTGTGCTGTGCTGTG 1981  
Db 1081 CATCAAGTGTGGCGCTATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1140  
QY 1982 AGCCAGCGCTTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2041  
Db 1141 AGCCAGAGAGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1200  
QY 2042 CTGACTCGGAAGCAGCGGAGTGTGACCCAGCTGTGA 2079  
Db 1201 CTGACTCGGAAGCAGTGGAGTGTGACTGAGCCCTGA 1238

RESULT 6

CR592189 1212 bp mRNA 11near HNC 21-JUL-2004  
LOCUS full-length cDNA clone CS0DM012Y015 of Fetal Liver of Homo sapiens  
(human).  
ACCESSION CR592189.1 GI:50472996  
VERSION CR592189.1  
KEYWORDS HTC; CNSLT cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1212)  
AUTHORS Li, W.B., Gruber, C., Jeese, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradey Avenue  
2 (bases 1 to 1212)  
REFERENCE 2 (bases 1 to 1212)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source 1..1212  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM012Y015"  
/tissue\_type="Fetal liver"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 38.3%; Score 905.4; DB 3; Length 1212;  
Best Local Similarity 99.3%; Pred. No. 6.2e-200;  
Matches 909; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1436 GCTTCCCGAGCGGAAACCGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 1495  
DB 298 GCTTCTCAGGCGGAAACCGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 357  
QY 1496 AGGCTCTCCGCAAGGCTGAGGCTGGGCTGAGGCTGCTGCTACCAAGGCGCTATC 1555  
DB 358 AGGCTCTCCGCAAGGCTGAGGCTGGGCTGAGGCTGCTGCTACCAAGGCGCTATC 417  
QY 1556 TTCCCTTCGGGAGCCCAACAGCGAGGAGAACAGAACAGATATTCCTGCTGCACTCT 1615  
DB 418 TTCCCTTCGGGAGCCCAACAGCGAGGAGAACAGAACAGATATTCCTGCTGCACTCT 477  
QY 1616 CCAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1675  
DB 478 CCAAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537  
QY 1676 CCGCTGAGATGAGCAAGATCTGTACCCGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 1735  
DB 538 CCGCTGAGATGAGCAAGATCTGTACCCGAGGCTGGGCTGGGCTGGGCTGGGCTGGGCT 597  
QY 1736 AACAGGCGGGGCTACTCCAGAGGCTGAGTCCCATATATCAGCATATATTCGCAATG 1795  
DB 598 AACAGGCGGGGCTACTCCAGAGGCTGAGTCCCATATATCAGCATATATTCGCAATG 657  
QY 1796 GCGCTGATTTCTATGAGAAACAGATCAAGCCCAAGATTTCTGCTGCTGCTGCTGCTGCT 1855  
DB 658 GCGCTGATTTCTATGAGAAACAGATCAAGCCCAAGATTTCTGCTGCTGCTGCTGCTGCT 717  
QY 1856 GTGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1915

DB 718 GTGCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777  
QY 1916 CTGAGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1975  
DB 778 CTGAGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 837  
QY 1976 CCGAGAGCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2035  
DB 838 CCGAGAGCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 897  
QY 2036 TAAAGTCACTCCGAGAGCCAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2095  
DB 898 TAAAGTCACTCCGAGAGCCAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957  
QY 2096 GCGAGGCTCCAGAGGCGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2155  
DB 958 GCGAGGCTCCAGAGGCGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017  
QY 2156 GACGTTTTCTTCTTGGGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2215  
DB 1018 GACGTTTTCTTCTTGGGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077  
QY 2216 TCTTCCAGTGGGCGGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2275  
DB 1078 TCTTCCAGTGGGCGGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137  
QY 2276 CATGTAATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2335  
DB 1138 CATGTAATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197  
QY 2336 TCTTAAATATATAA 2350  
DB 1198 TCTTAAATATATAA 1212

RESULT 7  
AL558100 846 bp mRNA 11near EST 02-APR-2004  
LOCUS full-length cDNA clone CS0DJ002YF16 5-PRIME, mRNA sequence.  
DEFINITION Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
ACCESSION AL558100  
VERSION AL558100.3 GI:46183500  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 846)  
AUTHORS Li, W.B., Gruber, C., Jeese, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 15, 2001 this sequence version replaced gi:31279899.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Creteil, CP 5706 - 91057 Evry cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
9952.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?b=CS0DJ002D08Q1&c=9952.f.  
FEATURES  
source 1..846  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ002YF16"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_1lb="Homo sapiens T CELLS (JURKAT CELL LINE) COT

10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."



|||||  
Db 779 CTATGAAACC--GATCAAGCCAGATTTCTGTGTCCTGATACCCCGAGGGGTGGCTTTG 836  
1865 ATGCTTCGACAGGCGACAGCCGTGTCTCTTTGTGTGTGAGACAGCATCTCTGTGAGAGC 1924  
Db 837 ATTCTGTGACAGGCGACCGCGGTGT-CTTTGTGTGTGAGAGCGGC-TCCTCGAGAGCG 894  
1925 CAGGTGGCGGCTGTGTG 1942  
Db 895 CCGTGGCGGCTGTGTG 912

RESULT 10  
BX406458 965 bp mRNA linear EST 01-MAY-2004  
LOCUS BX406458  
DEFINITION BX406458 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
ACCESSION CS0DM011Y04 5-PRIME, mRNA sequence.  
VERSION BX406458  
KEYWORDS BX406458.2 GI:46928188  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
AUTHORS Pull-length cDNA libraries and normalization  
TITLE Unpublished (2001)  
JOURNAL On May 13, 2003 this sequence version replaced gi.30644262.  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremlieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
This sequence belongs to sequence cluster 9952.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?cs=CS0AM011DB02Q1&c=9952.f.  
Location/Qualifiers  
1..965  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM011Y04"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo (dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

FEATURES  
source  
1..965  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM011Y04"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo (dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN  
Query Match 33.6%; Score 794; DB 5; Length 965;  
Best Local Similarity 93.1%; Pred. No. 5.5e-174;  
Matches 833; Conservative 16; Mismatches 44; Indels 2; Gaps 2;

Db 729 CTGGGCTCCCCAGGCGCTGCTCCCGTCATCTCCACAGAGTCCACCGTGGCCAG 788  
71 CGGACACTACTCTGAGGCTCCGCCCCACCTGCTGAGACCCAGGGTCCACCTTGCCAG 130  
789 GAGGTCAAGCCAGGGAATCATTAACAAGAGGAGTGCATGCGAGAGAGAGGCTGGCG 848  
Db 131 GAGGTCAAGCCAGGGAATCATTAACAAGAGGAGTGCATGCGAGAGAGAGGCTGGCG 190  
Gy 849 GACTGTGCATGCTGCTTCAGACCCAGAGTGGCAGCTTCACTGGGGGAGCTTGCTACT 908  
Db 191 GACTGTGCATGCTGCTTCAGACCCAGAGTGGCAGCTTCACTGGGGGAGCTTGCTACT 250

Gy 909 TCTGACAGCATCGGGCGGCGATCTCTGGGCAATTGTGGTCTCTCTCAGAGTAGACA 968  
Db 251 TCTGACAGCATCGGGCGGCGATCTCTGGGCAATTGTGGTCTCTCTCAGAGTAGACA 310  
Gy 969 GAGCGGCTGTACCCAGTGCAGAGTCAAGCTTGTCCGAGCGCTGTGATGTCTTTGACAA 1028  
Db 311 GAGCGGCTGTACCCAGTGCAGAGTCAAGCTTGTCCGAGCGCTGTGATGTCTTTGACAA 370  
Gy 1029 GAGCGAGGAGACGTGGGGGCTGTGTGTCTCTCTCGCGCTCCAAAGCCAGAGGAGCGAGCT 1088  
Db 371 GAGCGAGGAGACGTGGGGGCTGTGTGTCTCTCTCGCGCTCCAAAGCCAGAGGAGCGAGCT 430  
Gy 1089 CAGCTGCAGAGAGATGAGGCTTCTCAGAGGAGTGCACCCAGTCCAGAGTGCAGTGCAGAC 1148  
Db 431 CAGCTGCAGAGAGATGAGGCTTCTCAGAGGAGTGCACCCAGTCCAGAGTGCAGTGCAGAC 490  
Gy 1149 GCGGGGCGCCAAATGACAGTCCGCTTCTTGTGTGTGAGAGAGGAGCTGCCACAC 1208  
Db 491 GCGGGGCGCCAAATGACAGTCCGCTTCTTGTGTGTGAGAGAGGAGCTGCCACAC 550  
Gy 1209 CCAGAGGCTGCTGAGAGTCAATCCCGTGTGTGATGATGCGCCAGAGCGGTTCTTGGCGCG 1268  
Db 551 CCAGAGGCTGCTGAGAGTCAATCCCGTGTGTGATGATGCGCCAGAGCGGTTCTTGGCGCG 610  
Gy 1269 CATCTGCAGAGACTGTGTGCGCGCAGAGAGTGCACCGCATCTGTGGAGGCGCGAGA 1328  
Db 611 CATCTGCAGAGACTGTGTGCGCGCAGAGAGTGCACCGCATCTGTGGAGGCGCGAGA 669  
Gy 1329 CACCAAGTTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTGTGTGTGATGAGACACACTCTG 1388  
Db 670 CAMCAGCTTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTGTGTGATGAGAGACACACTCTG 728  
Gy 1389 TGGGGGATCCCTGCTCCCGGGGATCTGGGTGCTGACACCGGCCACGCTTCCGGAGCG 1448  
Db 729 TGGGGGATCCCTGCTCCCGGGGATCTGGGTGCTGACACCGGCCACGCTTCCGGAGCG 788  
Gy 1449 GAACCGGAGTCTGTCCCGATGAGTGCAGAGTGTGGCGGTGGCGCGAGCTTCTCCCA 1508  
Db 789 GAACCGGAGTCTGTCCCGATGAGTGCAGAGTGTGGCGGTGGCGCGAGCTTCTCCCA 848  
Gy 1509 CGGTCTGACCTGGGGGTGCGAGGCTGTGTGTACACAGGGGCTATCTTCCCTTGGGA 1568  
Db 849 CRGTCTGACCTGGGGGTGCGAGGCTGTGTGTACACAGGGGCTATCTTCCCTTGGGA 908  
Gy 1569 CCCCAGAGGAGAGAGAAACAGCAAGATATTCCTCGTCCACCTCCAGTCCC 1623  
Db 909 CCCCAGAGGAGAGAGAAACAGCAAGATATTCCTCGTCCACCTTCBAGSKCCC 963

RESULT 11  
CK231686  
LOCUS CK231686  
DEFINITION CK231686 1012 bp mRNA linear EST 09-DEC-2003  
ILLUMIGEN MCQ\_2776 Katze\_Mtld Macaca mulatta cDNA 5' similar to  
human HPN (hs.432750), mRNA sequence.  
ACCESSION CK231686  
VERSION CK231686.1 GI:39638044  
KEYWORDS EST.  
SOURCE Macaca mulatta (rhesus monkey)  
ORGANISM Macaca mulatta  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopithecoidea; Macaca.  
REFERENCE 1 (bases 1 to 1012)  
AUTHORS Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.  
TITLE Large-scale Rhesus Macaque cDNA Sequencing  
JOURNAL Unpublished (2003)  
COMMENT Contact: C. Magness  
Illumigen Biosciences Inc.  
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA  
Tel: 2063780400  
Fax: 2063780408  
Email: cmagness@illumigen.com  
Sequenced on 2003.11.05. 744 Q20 bases. Assemblies in contig w/ 2





|||||  
Db 1 CAGGCATCATTAACAAGAGCAGTGAATGGCCGAGAGAGAGGGTGGCCGACCTGTGCC 60  
Qy 858 ATGCTGCTCCAGACCCAGAGTGGGAGCTCTCACTGCGGGACCTCTACTTCTGACAGC 917  
Db 61 ATGCTGCTCCAGACCCAGAGTGGGAGCTCTCACTGCGGGACCTCTACTTCTGACAGC 120  
Qy 918 CATCGGGGGGCGCATCTCTGGGCGCATTTGATGGCTTCTCTCAAGAGTGAACAAGAGCGGCT 977  
Db 121 CATCGGGGGGCGCATCTCTGGGCGCATTTGATGGCTTCTCTCAAGAGTGAACAAGAGCGGCT 180  
Qy 978 GTACCCAGTGAAGGTCACTGCGGAGCTGCGGCTCAATGCTTTTGAACAAGAGAGG 1037  
Db 181 GTACCCAGTGAAGGTCACTGCGGAGCTGCGGCTCAATGCTTTTGAACAAGAGAGG 240  
Qy 1038 GACGTGGCGGCTGCTGTGCTCTCTGCGCTCCAGCCGAGGATGCGGACTCACTGGA 1097  
Db 241 GACGTGGCGGCTGCTGTGCTCTCTGCGCTCCAGCCAGCCCA-GGTAGCCGACTCACTGGA 299  
Qy 1098 GAGATGGGCTTCTCTCAAGGAGCTGACCCATCCGAGTGGACGTGAGAGCGGCGGCGC 1157  
Db 300 GAGATGGGCTTCTCTCAAGGAGCTGACCCATCCGAGTGGACGTGAGAGCGGCGGCGC 359  
Qy 1158 CAATGGACAGTGGGCTTCTCTGTGTGAAGAGGAGGAGCTGCCCAACCCAGAGCT 1217  
Db 360 CAATGGACAGTGGGCTTCTCTGTGTGAAGAGGAGGAGCTGCCCAACCCAGAGCT 419  
Qy 1218 GCTGAGAGTCACTCTCCGTGTGATGCGCCAGAGCGGCTTCTTGGCGGCGCATCTGCA 1277  
Db 420 GCTGAGAGTCACTCTCCGTGTGATGCGCCAGAGCGGCTTCTTGGCGGCGCATCTGCA 479  
Qy 1278 AGACTGTGGCGGAGAACTGCCGATGAGCGGATGTTGGAGGCGGAGCACACAGCTT 1337  
Db 480 AGACTGTGGCGGAGAACTGCCGATGAGCGGATGTTGGAGGCGGAGCACACAGCTT 539  
Qy 1338 GGGCGGCTGGCGGAGTCAAGCTTGTGTATGATGAGACACACTCTGTGGGAGATC 1397  
Db 540 GGGCGGCTGGCGGAGTCAAGCTTGTGTATGATGAGACACACTCTGTGGGAGATC 599  
Qy 1398 CTTGCTCTCCGGGAGTGGGTGCTGACAGCGGCGGCACTGCTCCGAGGCGGAACTGGGT 1457  
Db 600 CTTGCTCTCCGGGAGTGGGTGCTGACAGCGGCGGCACTGCTCCGAGGCGGAACTGGGT 659  
Qy 1458 CTTGCTCCGATGGGAGATG-TTTGGCGGCTGGCGGAGCGGCTCT-CCGCAAGGCTGTG 1515  
Db 660 CTTGCTCCGATGGGAGATGTTTGGCGGCTGGCGGAGCGGCTCTCTCTCCGCAAGGCTGTG 719  
Qy 1516 CAGCT-GGGGAGTGAAGCTGTGTCTACACAGGAGGCTATCTTCC-TTTGGGAGCCCA 1573  
Db 720 CAGCTGGGAGGAGTGAAGCTGTGTCTACACAGGAGGCTATCTTCC-TTTGGGAGCCCA 779  
Qy 1574 ACACGAGAGAAACAAGACATATGCGCTGTGCTCACTCTCCAGTCCCTGCGGCTCA 1633  
Db 780 ACACGAGAGAAACAAGACATATGCGCTGTGCTCACTCTCTCACTCTGCGGCTCAAG 839  
Qy 1634 CAGATATCATCAGCTGTGTGCTCCGAGCTGCGGCGGAGAGGCT 1677  
Db 840 AATTAATCA-----GCTGTGTGCTCTCA-CTGCGGCGGAGAGGCT 877

RESULT 13  
BX422288 961 bp mRNA linear EST 01-MAY-2004  
LOCUS BX422288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
DEFINITION BX422288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone  
ACCESSION BX422288  
VERSION BX422288.2 GI:46933436  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 961)

AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 13, 2003 this sequence version replaced gi:30655303.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremlieux, CP 5706 - 91057 Evry cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
This sequence belongs to sequence cluster 9952.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdnas-CS0DM003BD05QPL4c=9952.f.  
Location/Qualifiers  
1. 961  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DM003YGI0"  
/tissue\_type="FETAL LIVER"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL LIVER"  
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN  
Query Match 31.9%; Score 753; DB 5; Length 961;  
Best Local Similarity 97.4%; Pred. No. 2e-164;  
Matches 817; Conservative 1; Mismatches 16; Indels 5; Gaps 5;  
Qy 736 CCCCAGAGCCCTGCTCCCTCCATCTCTCAAGAGTCCACCTGAGCCAGAGGTCA 795  
Db 78 CTTGAGGCTCCGCCCCACCTGTGAGCCCA-GGCTCCACCTGCGCCAGAGGTCA 136  
Qy 796 GCCAGGAAATCAATTAACAAGAGGAGTGAATGGCGGAGAAAGAGGTGGCCGAGCTGTG 855  
Db 137 GCCAGGAAATCAATTAACAAGAGGAGTGAATGGCGGAGAAAGAGGTGGCGGAGCTGTG 196  
Qy 856 CCATGCTGCTCCAGACCCAGAGTGGGAGCTCTCACTGCGGAGGAGCCCTGACTTCTGACA 915  
Db 197 CCATGCTGCTCCAGACCCAGAGTGGGAGCTCTCTCACTGCGGAGGAGCCCTGACTTCTGACA 256  
Qy 916 GCCATCGGAGCGGATCTGGGCAATTTGTGCTGTTCTCTCAAGAGTGAACAAGAGCGG 975  
Db 257 GCCATCGGAGCGGATCTGGGCAATTTGTGCTGTTCTCTCAAGAGTGAACAAGAGCGG 316  
Qy 976 CTGTATCCAGTGAAGTCACTCTGCGGAGAGCTGCGGCTCATGTTGTTTGAACAAGAGGAA 1015  
Db 317 CTGTATCCAGTGAAGTCACTCTGCGGAGAGCTGCGGCTCATGTTGTTTGAACAAGAGGAA 376  
Qy 1036 GGGAGCGTGGGAGCTGTGCTCTCTGCGGCTCAAGCGGAGGAGCGGAGCTCAGCTGC 1095  
Db 377 GGGAGCGTGGGAGCTGTGCTCTCTGCGGCTCAAGCGGAGGAGGAGCTCAGCTGC 436  
Qy 1096 GAGAGAGATGGGCTTCTCTCAAGGAGCTGACCCATCCAGGCTGAGCGTGAACAAGCGGAGC 1155  
Db 437 GAGAGAGATGGGCTTCTCTCAAGGAGCTGACCCATCCAGGCTGAGCGTGAACAAGCGGAGC 496  
Qy 1156 GCCAATGGCAAGTGGGCTTCTCTGTGTGAAGAGGAGGAGCTGCCCAACCAAGAGG 1215  
Db 497 GCCAATGGCAAGTGGGCTTCTCTGTGTGAAGAGGAGGAGGCTGCCCAACCAAGAGG 556  
Qy 1216 CTGCTGAAGTCACTCTCCGTGTGATGCGCCAGAGAGCGGTTCTTGGCGGCACTGTC 1275  
Db 557 CTGCTGAAGTCACTCTCCGTGTGATGCGCCAGAGAGCGGTTCTTGGCGGCACTGTC 616  
Qy 1276 CAAGACTGTGGCGGAGGAAGCTGCCGCTGAGACCGACTGTGGAGGCGGAGCACACGAGC 1335

| Accession | Sequence   | Position |
|-----------|--|----------|
| Db        | 617 CAAGA CTGTG CCGCAGAAAGCTGGACCCGTGACCGCATGTGTGGAGCCCGGGA CACGAC   | 676      |
| Qy        | 1336 TTGGGCGCGGTGCGCGGTGGCAAGTACGCTTGCGCTATGATGTGAGACACACTCTGTGGGGGA | 1399     |
| Db        | 677 TTGGGCGCGGTGCGCGGTGGCAAGTACGCTTGCGCTATGATGTGAGACACACTCTGTGGGGGA  | 736      |
| Qy        | 1396 TCCTGCTCTCCGGGGA CTGGGTGTGACAGCCCGCCCACTGCTTCCGGAGCGGAACCGG     | 1455     |
| Db        | 737 T-CTGTGCTCTCCGGGGA CTGGGTGTGACAGCCCGCCCACTGCTTCCGGAGCGGA- CGG    | 794      |
| Qy        | 1456 GTCTGTCTCCGATGGCAGATGTTTGCCGATCCCGTGAGCCAGAGCCCTCCCAAGGCTG      | 1515     |
| Db        | 795 GTCTGTCTCCGATGGCAGATGTTTGCCGATCCCGTGAGCCAGAGCCCTCCCAAGGCTG       | 854      |
| Qy        | 1516 CAGCTGGGGGAGCAGGCTGTGTCTACACACGGGGCTATCTTCCTTTCCGGAGACCCCA      | 1574     |
| Db        | 855 CAGCTGGGGGAG- CAGGCTGTGTCTA- CACGGGGGCTATCTTCCTTTCCGGAGACCCCA    | 911      |

| RESULT 14  |                       |             |      |        |                 |
|------------|-----------------------|-------------|------|--------|-----------------|
| LOCUS      | BX427780/c            |             |      |        |                 |
| DEFINITION | BX427780              | 874 bp      | mRNA | linear | EST 03-MAY-2004 |
| ACCESSION  | BX427780              |             |      |        |                 |
| VERSION    | CS00DM012X015.3-PRIME |             |      |        |                 |
| KEYWORDS   | BX427780              |             |      |        |                 |
| SOURCE     | BX427780.2            | GI:46959112 |      |        |                 |
| ORGANISM   | Homo sapiens          |             |      |        |                 |
|            | Homo sapiens (human)  |             |      |        |                 |
|            | Homo sapiens          |             |      |        |                 |

REFERENCE 1 (bases 1 to 874)  
AUTHORS I. W. B., Gruber, C., Jessup, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 15, 2003 this sequence version replaced gi:30770691.

Genoscope - Centre National de Séquençage  
2 rue Gaston Crémieux, CP 5706 - 91057 Evry cedex - FRANCE  
Email: [seqret@genoscope.cns.fr](mailto:seqret@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Left strand cDNA was primed with a NotI-ciligo (dT) primer. Five primed  
ends enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

This sequence belongs to sequence cluster 9952.f  
For more information about this cluster, see  
<http://www.genoscope.cns.fr/cdna/s=CS0AM012AH08NP1&c=9952.f>

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM012Y015"
/rnause_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6, 1st strand cDNA
was primed with a NotI-oligo(dT) primer, five prime end
enriched, double-strand cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```

Query Match 31.6%; Score 747; DB 5; Length 874;  
Best Local Similarity 95.6%; Pred. No. 4,99-163;  
Matches 813; Conservative 0; Mismatches 20; Indels 9; Gaps 5;

|   |      |  |      |
|---|------|--|------|
| Oy  | 1525 | GTGAGAGGTGGTGGCTACGACAGGGGGGCTATCTTCCCTTGGGAGCCCAACAGGAGAG         | 158  |
| Db  | 818  | GTGAGAGGTGT- <u>GTCTACACAGGG</u> ---CTATCTCTCTTTCGGACCCCAACAGGAGAG | 763  |
| Oy  | 1585 | AACGCAACGATATTTGGCCCTGGTCCACCTCTCCAGTCCCTCGCCCTCACAAATACATC        | 1644 |
| Db  | 762  | AACGCAACGATATTTGGCCCTGGTCCACCTCTCCAGTCCCTCGCCCTCACAAATACATC        | 703  |
| Oy  | 1645 | CAGCCTGTGTGCTTCCAGCTGCGGCGAGGCCCTGTGTGATGGCAAGATTTGATCCGTG         | 1704 |
| Db  | 702  | CAGCCTGTGTGCTTCCAGCTGCGGCGAGGCCCTGTGTGATGGCAAGATTTGATCCGTG         | 643  |
| Oy  | 1705 | ACGGGCTGGGGCAACAGCAGTACTATAGGCCAAAGGCCGGGGATACCTCCAGAGGCTCGA       | 1764 |
| Db  | 642  | ACGGGCTGGGGCAACAGCAGTACTATAGGCCAAAGGCCGGGGATACCTCCAGAGGCTCGA       | 583  |
| Oy  | 1765 | GTCCCCATTAATCAGCAATGATGTCTGCATATGGCGCTGACTTCTATGGAAACCAATCAG       | 1824 |
| Db  | 582  | GTCCCCATTAATCAGCAATGATGTCTGCATATGGCGCTGACTTCTATGGAAACCAATCAG       | 523  |
| Oy  | 1825 | CCCAAGATGTTCTGTGTGGCTACCCCGAGGGTGCAATGATGCTTGCACAGGGCGACAGC        | 1884 |
| Db  | 522  | CCCAAGATGTTCTGTGTGGCTACCCCGAGGGTGCAATGATGCTTGCACAGGGCGACAGC        | 463  |
| Oy  | 1885 | GGTGTGTCCCTTTGTGTGTGAGAGACAGATCTCTGGAGCGCAAGTTGGCGGCTGTGTGGC       | 1944 |
| Db  | 462  | GGTGTGTCCCTTTGTGTGTGAGAGACAGATCTCTGGAGCGCAAGTTGGCGGCTGTGTGGC       | 403  |
| Oy  | 1945 | ATTGTGATTTGGGGCACATGGCTGTGGCCCTGGCCGAGAAGCCAGGCGTCTACACAAAGTC      | 2004 |
| Db  | 402  | ATTGTGATTTGGGGCACATGGCTGTGGCCCTGGCCGAGAAGCCAGGCGTCTACACAAAGTC      | 343  |
| Oy  | 2005 | AGTGACTTCCGGGAGTGGATCTTTCAGAGCCATAAAGACTCACTCGAAGCCAGCGGATG        | 2064 |
| Db  | 342  | AGTGACTTCCGGGAGTGGATCTTTCAGAGCCATAAAGACTCACTCGAAGCCAGCGGATG        | 283  |
| Oy  | 2065 | GTGACCCAGCTCTGACCGGTGGCTTCTGTGCTGCGCAGCTTCAGAGGCCAGAGTGAATCC       | 2124 |
| Db  | 282  | GTGACCCAGCTCTGACCGGTGGCTTCTGTGCTGCGCAGCTTCAGAGGCCAGAGTGAATCC       | 224  |
| Oy  | 2125 | GGTGTGTGGGATCCACGCTGGGGCCGAGGATGGGAAGTTTTCTTCTGTGGGCCGGGTCCACA     | 2184 |
| Db  | 223  | GGTGTGTGGGATCCACGCTGGGGCCGAGGATGGGAAGTTTTCTTCTGTGGGCCGGGTCCACA     | 164  |
| Oy  | 2185 | GGTCCAAAGACACCCCTCCCTCAAGGATCTTCTTCCACAGTGGCGGGGCCACTCAGGCC        | 2244 |
| Db  | 163  | GGTCCAAAGACACCCCTCCCTCAAGGATCTTCTTCCACAGTGGCGGGGCCACTCAGGCC        | 104  |
| Oy  | 2245 | CGAGACCAACCCCACTCACCTCTCTGACCCCCCATGTAAATATTGTTCTGTCTGTGGAC        | 2304 |
| Db  | 103  | CGAGACCAACCCCACTCACCTCTCTGACCCCCCATGTAAATATTGTTCTGTCTGTGGAC        | 44   |
| Oy  | 2305 | TCTGTGTCTAG 2314   |      |
| Db  | 43   | ACCTGTCTTG 34  |      |
| RESULT 15   |      |  |      |
| B1752774  |      |  |      |
| LOCUS   |      |  |      |
| DEFINITION  |      |  |      |
| 60302181.F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5192427 5', |      |  |      |
| mRNA sequence.  |      |  |      |
| ACCESSION   |      |  |      |
| B1752774  |      |  |      |
| VERSION   |      |  |      |
| B1752774.1 GI:15744352  |      |  |      |
| KEYWORDS  |      |  |      |
| EST.  |      |  |      |
| SOURCE  |      |  |      |
| ORGANISM  |      |  |      |
| Homo sapiens (human)  |      |  |      |
| Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |      |  |      |
| Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.         |      |  |      |
| REFERENCE   |      |  |      |
| 1 NIH-MGC http://mgc.nci.nih.gov/.                                |      |  |      |
| 2 National Institutes of Health, Mammalian Gene Collection (MGC)  |      |  |      |
| JOURNAL   |      |  |      |
| Published (1999)  |      |  |      |



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 07:27:22 ; Search time 4014 Seconds

(without alignments)  
3824.668 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363  
Sequence: 1 cggagccgcttc...cgga.....taataagatggtttgatt 2363

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapcut 1.0

Searched: 7316285 seqs, 3211459403 residues

Total number of hits satisfying chosen parameters: 14632570

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Published Applications NA:\*

1: /cgn2\_6/ptodate/1/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodate/1/pubna/US06\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodate/1/pubna/US05\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodate/1/pubna/US04\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodate/1/pubna/US03\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodate/1/pubna/US02\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodate/1/pubna/US01\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodate/1/pubna/US00\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodate/1/pubna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodate/1/pubna/US08\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodate/1/pubna/US07\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodate/1/pubna/US06\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodate/1/pubna/US05\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodate/1/pubna/US04\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodate/1/pubna/US03\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodate/1/pubna/US02\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodate/1/pubna/US01\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodate/1/pubna/US00\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodate/1/pubna/US09\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodate/1/pubna/US08\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodate/1/pubna/US07\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodate/1/pubna/US06\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodate/1/pubna/US05\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodate/1/pubna/US04\_PUBCOMB.seq.\*  
25: /cgn2\_6/ptodate/1/pubna/US03\_PUBCOMB.seq.\*  
26: /cgn2\_6/ptodate/1/pubna/US02\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description         |
|------------|--------|-------------|--------|----|---------------------|
| 1          | 2363   | 100.0       | 2363   | 9  | US-09-880-107-3690  |
| 2          | 2363   | 100.0       | 2363   | 11 | US-09-968-0072A-436 |
| 3          | 2363   | 100.0       | 2363   | 17 | US-10-099-322-39    |
| 4          | 2363   | 100.0       | 2363   | 21 | US-10-044-564-39    |
| 5          | 2363   | 100.0       | 2363   | 17 | US-10-843-641A-6906 |
| 6          | 1593.6 | 67.4        | 1783   | 9  | US-09-919-048-188   |
| 7          | 1593.6 | 67.4        | 1783   | 10 | US-09-776-191-67    |

|    |        |      |       |    |                      |                      |
|----|--------|------|-------|----|----------------------|----------------------|
| 8  | 1593.6 | 67.4 | 1783  | 14 | US-10-102-283-188    | Sequence 188, App    |
| 9  | 1593.6 | 67.4 | 1783  | 14 | US-10-073-060-1      | Sequence 1, Appl     |
| 10 | 1593.6 | 67.4 | 1783  | 15 | US-10-135-795-188    | Sequence 188, App    |
| 11 | 1593.6 | 67.4 | 1783  | 15 | US-10-205-823-177    | Sequence 177, App    |
| 12 | 1593.6 | 67.4 | 1783  | 16 | US-10-210-120-1      | Sequence 1, Appl     |
| 13 | 1593.6 | 67.4 | 1783  | 17 | US-10-295-027-1160   | Sequence 1160, App   |
| 14 | 1593.6 | 67.4 | 1783  | 17 | US-10-156-214A-34    | Sequence 34, Appl    |
| 15 | 1593.6 | 67.4 | 1783  | 19 | US-10-652-993-188    | Sequence 188, App    |
| 16 | 1593.6 | 67.4 | 1783  | 21 | US-10-956-157-792    | Sequence 792, App    |
| 17 | 1593.6 | 67.4 | 1783  | 22 | US-10-909-035-1      | Sequence 1, Appl     |
| 18 | 1482   | 62.7 | 1482  | 21 | US-10-956-157-792    | Sequence 1, Appl     |
| 19 | 1400   | 59.2 | 1400  | 21 | US-10-956-157-792    | Sequence 1, Appl     |
| 20 | 1269   | 53.7 | 1615  | 9  | US-09-820-002-1      | Sequence 6027, App   |
| 21 | 1269   | 53.7 | 1615  | 15 | US-10-274-031-1      | Sequence 1, Appl     |
| 22 | 1254   | 53.1 | 1254  | 18 | US-10-235-789-1      | Sequence 1, Appl     |
| 23 | 1097.8 | 46.5 | 7033  | 19 | US-10-678-816-5      | Sequence 5, Appl     |
| 24 | 1097   | 46.4 | 12242 | 17 | US-10-678-816-6      | Sequence 6, Appl     |
| 25 | 858.4  | 36.3 | 2175  | 19 | US-10-108-260A-559   | Sequence 559, App    |
| 26 | 778.2  | 32.9 | 11924 | 19 | US-10-678-816-7      | Sequence 7, Appl     |
| 27 | 402.8  | 17.0 | 493   | 10 | US-09-918-995-32762  | Sequence 32762, App  |
| 28 | 380.2  | 16.1 | 497   | 19 | US-10-250-889-38     | Sequence 38, Appl    |
| 29 | 374.8  | 15.9 | 494   | 10 | US-09-918-995-32775  | Sequence 3, Appl     |
| 30 | 325    | 13.8 | 21784 | 15 | US-09-820-002-3      | Sequence 3, Appl     |
| 31 | 325    | 13.8 | 21784 | 15 | US-10-274-031-3      | Sequence 3, Appl     |
| 32 | 321.8  | 13.6 | 2547  | 17 | US-10-104-047-819    | Sequence 819, App    |
| 33 | 279    | 11.8 | 279   | 18 | US-10-424-599-133667 | Sequence 133667, App |
| 34 | 249.8  | 10.6 | 266   | 9  | US-09-964-824A-462   | Sequence 462, App    |
| 35 | 249.8  | 10.6 | 266   | 9  | US-09-880-107-1812   | Sequence 1812, App   |
| 36 | 249.8  | 10.6 | 266   | 21 | US-10-843-641A-5765  | Sequence 5765, App   |
| 37 | 192    | 8.1  | 601   | 9  | US-09-820-002-15     | Sequence 15, Appl    |
| 38 | 192    | 8.1  | 601   | 15 | US-10-274-031-15     | Sequence 15, Appl    |
| 39 | 191.4  | 8.1  | 568   | 16 | US-10-029-386-2340   | Sequence 2340, App   |
| 40 | 191    | 8.1  | 601   | 9  | US-09-820-002-16     | Sequence 16, Appl    |
| 41 | 191    | 8.1  | 601   | 15 | US-10-274-031-16     | Sequence 16, Appl    |
| 42 | 190    | 8.0  | 197   | 16 | US-10-029-386-16040  | Sequence 16040, App  |
| 43 | 185.6  | 7.9  | 379   | 9  | US-09-960-352-7479   | Sequence 7479, App   |
| 44 | 177.2  | 7.5  | 357   | 9  | US-09-960-352-4196   | Sequence 4196, App   |
| 45 | 169.4  | 7.2  | 601   | 9  | US-09-820-002-14     | Sequence 14, Appl    |

#### ALIGNMENTS

RESULT 1  
US-09-880-107-3690  
; Sequence 3690, Application US/09880107  
; Patent No. US20020142981A1  
GENERAL INFORMATION:  
; APPLICANT: Horne, Darcie T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3690  
; LENGTH: 2363  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X07732  
US-09-880-107-3690  
Query Match 100.0%; Score 2363; DB 9; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

|    |      |   |      |
|----|------|---|------|
| QY | 1081 | CCCCGACTCAGCTCCGAGAGAAATGGGCTTCCTCAAGGGCACTGACCCACTCCGAACTGGAC    | 1140 |
| Db | 1081 | GCCGACTCAGACTCCGAGAGATGGGCTTCCTCAAGGGCACTGACCCACTCCGAACTGGAC      | 1140 |
| QY | 1141 | GTGGAAACGGGGGCGCCAAATGAGCAGCTGGGGCTTCCTCTGTGTGAGACGAGGGGAGGCTG    | 1200 |
| Db | 1141 | GTGGAAACGGGGGCGCCAAATGAGCAGCTGGGGCTTCCTCTGTGTGAGACGAGGGGAGGCTG    | 1200 |
| QY | 1201 | CCCCACACCCAGAGGCTGTGGAGTCACTCCGTGTGTGATTGCCCCAGAGGCCGTTTC         | 1266 |
| Db | 1201 | CCCCACACCCAGAGGCTGTGGAGTCACTCCGTGTGTGATTGCCCCAGAGGCCGTTTC         | 1266 |
| QY | 1261 | TTTGGCCCGCATCTTCCCAAGACTGTGTGGCCGAGAAAGCTGCCCTGTGACCCGATGTGGGA    | 1320 |
| Db | 1261 | TTTGGCCCGCATCTTCCCAAGACTGTGTGGCCGAGAAAGCTGCCCTGTGACCCGATGTGGGA    | 1320 |
| QY | 1321 | GGCCGGGACACACACTTGTGGCCGGTGGCCGGTGGACAGTCAAGCTTCCTCATATGATGAGCA   | 1380 |
| Db | 1321 | GGCCGGGACACACACTTGTGGCCGGTGGCCGGTGGACAGTCAAGCTTCCTCATATGATGAGCA   | 1380 |
| QY | 1381 | CACCTCTGTGTGGGGATCCCTGTCTCTCCGGGAGCTGGGTGTGACAGGCCGCCCATGTCTTC    | 1440 |
| Db | 1381 | CACCTCTGTGTGGGGATCCCTGTCTCTCCGGGAGCTGGGTGTGACAGGCCGCCCATGTCTTC    | 1440 |
| QY | 1441 | CCGAGAGCGGAAACCGGGCTCTGTCCCGATGTGCGAGTGTTCGCCGGTGTCCCTGTGGCCAGGCC | 1500 |
| Db | 1441 | CCGAGAGCGGAAACCGGGCTCTGTCCCGATGTGCGAGTGTTCGCCGGTGTCCCTGTGGCCAGGCC | 1500 |
| QY | 1501 | TCTCCCAACGGTCTGACAGCTGGGGGTGACAGCTGTGGTCTACACAGGGGGGCTATCTTCCC    | 1560 |
| Db | 1501 | TCTCCCAACGGTCTGACAGCTGGGGGTGACAGCTGTGGTCTACACAGGGGGGCTATCTTCCC    | 1560 |
| QY | 1561 | TTTTGGGAGACCCCAACAGCGAGGAGAACAGACGATATTGCGCTGTGTCACACTTCCAGT      | 1620 |
| Db | 1561 | TTTTGGGAGACCCCAACAGCGAGGAGAACAGACGATATTGCGCTGTGTCACACTTCCAGT      | 1620 |
| QY | 1621 | CCCTGTCCCTCAACAGAAATACATCAGCTGTGTGTCTCCAGCTGCCCGCCAGGCCCTG        | 1680 |
| Db | 1621 | CCCTGTCCCTCAACAGAAATACATCAGCTGTGTGTCTCCAGCTGCCCGCCAGGCCCTG        | 1680 |
| QY | 1681 | GTGATGTGGCAAGATCTGTACCGGTACGGGGCTGGGGGCAACAGCAGTACTATGTGGCCACAG   | 1740 |
| Db | 1681 | GTGATGTGGCAAGATCTGTACCGGTACGGGGCTGGGGGCAACAGCAGTACTATGTGGCCACAG   | 1740 |
| QY | 1741 | GCCGGGGTACTCCAGAGAGGCTCGAGTCCGCCATATCAGCAATGATGTCTGCAATGAGCGCT    | 1800 |
| Db | 1741 | GCCGGGGTACTCCAGAGAGGCTCGAGTCCGCCATATCAGCAATGATGTCTGCAATGAGCGCT    | 1800 |
| QY | 1801 | GACTTCTATGTGAAACCAAGATCAAGCCCAAGATGTCTGTGTGTGTGAGGACAGCATCTCCGG   | 1860 |
| Db | 1801 | GACTTCTATGTGAAACCAAGATCAAGCCCAAGATGTCTGTGTGTGTGAGGACAGCATCTCCGG   | 1860 |
| QY | 1861 | ATTGATGTGCGCCAGGGGCGACAGGGGTGTCCCTTGTGTGTGAGGACAGCATCTCCGG        | 1920 |
| Db | 1861 | ATTGATGTGCGCCAGGGGCGACAGGGGTGTCCCTTGTGTGTGAGGACAGCATCTCCGG        | 1920 |
| QY | 1921 | ACGCGCAGTTTGGCGGCTGTGTGTGATTTGTGATTTGGGGCACTGGCTGTGGCTTGGCCAG     | 1980 |
| Db | 1921 | ACGCGCAGTTTGGCGGCTGTGTGTGATTTGTGATTTGGGGCACTGGCTGTGGCTTGGCCAG     | 1980 |
| QY | 1981 | AAGCCAGGCGTCTACACAAAGTCAGTCACTTCGGGAGTGGATCTTCCAGGCCATTAAG        | 2040 |
| Db | 1981 | AAGCCAGGCGTCTACACAAAGTCAGTCACTTCGGGAGTGGATCTTCCAGGCCATTAAG        | 2040 |
| QY | 2041 | ACTCACTCCGAAGCCAGCGGATGTGTACCCAGCTTGACCCGGTGTCTTCGCTGTGGCA        | 2100 |
| Db | 2041 | ACTCACTCCGAAGCCAGCGGATGTGTACCCAGCTTGACCCGGTGTCTTCGCTGTGGCA        | 2100 |
| QY | 2101 | GCCTCAGAGGCGGAGGTGATCCGGGTGTGGATTCACGCTGGGCCGAGATGGAGCT           | 2160 |
| Db | 2101 | GCCTCAGAGGCGGAGGTGATCCGGGTGTGGATTCACGCTGGGCCGAGATGGAGCT           | 2160 |
| QY | 2161 | TTTTTCTTGTGGCGCGGTCCACAGGTCCAAAGCACCTTCCTCCAGGGTCTCTCTTC          | 2220 |



|    |      |   |      |
|----|------|---|------|
| Db | 2161 | TTTTCTTCTTGAGCCCGGCTCCACAGSTCCAGAGACACCTCTCCCTCCAGAGGTCCTCTCTTC | 2220 |
| Oy | 2221 | CACAGTGGCGGGCCACCTCAGCCCCGAGAGACCACTCACTCCTCTGACCCCGCANGT       | 2280 |
| Db | 2221 | CACAGTGGCGGGCCACCTCAGCCCCGAGAGACCACTCACTCCTCTGACCCCGCANGT       | 2280 |
| Oy | 2281 | AAATATTGTCGCTGCTGGGACGCTGCTAGGTGCCCCGTATGATGGAGTGTCTTT          | 2340 |
| Db | 2281 | AAATATTGTCGCTGCTGGGACGCTGCTAGGTGCCCCGTATGATGGAGTGTCTTT          | 2340 |
| Oy | 2341 | AAATATAAAGATGCTTTTGATT  | 2363 |
| Db | 2341 | AAATATAAAGATGCTTTTGATT  | 2363 |

RESULT 2  
US-09-968-007A-436  
; Sequence 436, Application US/09968007A  
; Publication No. US20040115625A1

```

APPLICANT: Ebner, Reinhard
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signa
TITLE OF INVENTION: Cancer Gene Sets
FILE REFERENCE: 689290-71
CURRENT APPLICATION NUMBER: US/09/968,007A
CURRENT FILING DATE: 2001-10-02
PRIORITY APPLICATION NUMBER: US/60/237,172
PRIORITY FILING DATE: 2000-10-02
PRIORITY APPLICATION NUMBER: US/60/237,173
PRIORITY FILING DATE: 2000-10-02
PRIORITY APPLICATION NUMBER: US/60/237,278
PRIORITY FILING DATE: 2000-10-02
PRIORITY APPLICATION NUMBER: US/60/237,294
PRIORITY FILING DATE: 2000-10-02
PRIORITY APPLICATION NUMBER: US/60/237,295
PRIORITY FILING DATE: 2000-10-02
PRIORITY APPLICATION NUMBER: US/60/237,316
PRIORITY FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 1001
SOFTWARE: PatentIn version 3.0
SEQ ID NO 436
LENGTH: 2363
TYPE: DNA
ORGANISM: Homo sapiens
US-09-968-007A-436

```

|                       |         |              |        |              |
|-----------------------|---------|--------------|--------|--------------|
| Query Match           | 100.0%; | Score 2363;  | DB 11; | Length 2363; |
| Best Local Similarity | 100.0%; | Pred. No. 0; |        |              |

|    |     |                             |                               |                          |                     |     |
|----|-----|-----------------------------|-------------------------------|--------------------------|---------------------|-----|
| Qy | 1   | TCGAGCCCGCTTTTCCAGAGGACCTTA | CTGAGGGGCCCA                  | CAGGTGAGGCA              | CGCTGGCTTAC         | 60  |
| Db | 1   | TCGAGCCCGCTTTCCAGAGGACCTTA  | CTGAGGGGCCCA                  | CAGGTGAGGCA              | CGCTGGCTTAC         | 60  |
| Qy | 61  | AGGACCCACGAGCA              | CCGCGCTGCGCTCC                | CGACGGCCCGCGCTGCGGGGGGCA | CATTGCTCC           | 120 |
| Db | 61  | AGGCCCCACGCGCACCGGCTT       | GTGCTTCCAGGCCGCGCGCTGCGGGGGCA | CACTAGCTTC               |                     | 120 |
| Qy | 121 | TGCCACGAGCCTGAGAGACTGAC     | CCCCGACCTCCAGGCTCCGCCCCCA     | CTGAC                    |                     | 180 |
| Db | 121 | TGCCACGAGCCTGAGAGACTGAC     | CCCCGACCTCCAGGCTCCGCCCCCA     | CTGAC                    |                     | 180 |
| Qy | 181 | TGAGCCCCAGGGTAA             | GACAGAGGGCCCCCA               | CACTACAGTTTC             | AGCCCTTGAGGACAGAGGG | 240 |
| Db | 181 | TGAGCCCCAGGGTAA             | GACAGAGGGCCCCCA               | CACTACAGTTTC             | AGCCCTTGAGGACAGAGGG | 240 |
| Qy | 241 | TTCCCTCATCCCGCCCA           | CCAGCCTTAATGCGCCCA            | CTCTTAATGAGAGGGTT        | CTCTGAGGAGCC        | 300 |
| Db | 241 | TTCCCTCATCCCGCCCA           | CCAGCCTTAATGCGCCCA            | CTCTTAATGAGAGGGTT        | CTCTGAGGAGCC        | 300 |
| Qy | 301 | TGAAGAGGGGCA                | CTATAGAGCTTCC                 | CCCAAGCACTAGAGTGTCT      | GTCTGTCTTCTTCTT     | 360 |

|    |      |  |      |
|----|------|--|------|
| Db | 301  | TGAAGAGGGGCACTTAAGACGTCTTCGCCAAGACCTAGGTGTCTGTCTGTCTTCTTCTT        | 360  |
| Qy | 361  | CAGACTCAGCGGTGTGACCCCAAGTCTTTTCTCTCCAGACCCAGAGATTTCCAGCCCTCAG      | 420  |
| Db | 361  | CAGACTCAGCGGTGTGACCCCAAGTCTTTTCTCTCCAGACCCAGAGATTTCCAGCCCTCAG      | 420  |
| Qy | 421  | GCCCCCTCTCCTCTCACTAGAGAGTCTGTGGCCCCCAATTCCTCTCTTTCCCAAGACTTA       | 480  |
| Db | 421  | GCCCCCTCTCCTCTCACTAGAGAGTCTGTGGCCCCCAATTCCTCTCTTTCCCAAGACTTA       | 480  |
| Qy | 481  | TGATTTACAGGTCTCAAGTGTCTCTCTCCCTCAAAACCGGAGATCTCAAGTCCCTGTCTCAC     | 540  |
| Db | 481  | TGATTTACAGGTCTCAAGTGTCTCTCTCCCTCAAAACCGGAGATCTCAAGTCCCTGTCTCAC     | 540  |
| Qy | 541  | CAGGCTCAGGATATGGGGGTCCCAATCCCTGTGAAATCCAGAGGTCCTCCCGCTGTCTGTCA     | 600  |
| Db | 541  | CAGGCTCAGGATATGGGGGTCCCAATCCCTGTGAAATCCAGAGGTCCTCCCGCTGTCTGTCA     | 600  |
| Qy | 601  | GACACTGACCCCAATCTTTGAAACCCAGCCCAATCTGCGTCCGTGTATCAAGGCGTGTCTG      | 660  |
| Db | 601  | GACACTGACCCCAATCTTTGAAACCCAGCCCAATCTGCGTCCGTGTATCAAGGCGTGTCTG      | 660  |
| Qy | 661  | CCAAGGCCCAATCCCTTCAAGGCTCTCCCTGAGATGAGACGCTTGAGACTGTGGGGGGCCAGAGCT | 720  |
| Db | 661  | CCAAGGCCCAATCCCTTCAAGGCTCTCCCTGAGATGAGACGCTTGAGACTGTGGGGGGCCAGAGCT | 720  |
| Qy | 721  | GGGCTGTGGGCTTGGGCTCCCCAGGCGCTGTCTCCCGCTCATCTCTCTCAAGATTCACACC      | 780  |
| Db | 721  | GGGCTGTGGGCTTGGGCTCCCCAGGCGCTGTCTCCCGCTCATCTCTCTCAAGATTCACACC      | 780  |
| Qy | 781  | TGGCCCAAGAGGTCAAGCCAGGGAATCATTTAACAGAGCAATGACATGGCGCAGAAAGAG       | 840  |
| Db | 781  | TGGCCCAAGAGGTCAAGCCAGGGAATCATTTAACAGAGCAATGACATGGCGCAGAAAGAG       | 840  |
| Qy | 841  | GGTGGCCGGAATGTGGCCATGTCTCTCCAGACCCAAAGTGAGAGCTCTCACTGTCCGGGAGCC    | 900  |
| Db | 841  | GGTGGCCGGAATGTGGCCATGTCTCTCCAGACCCAAAGTGAGAGCTCTCACTGTCCGGGAGCC    | 900  |
| Qy | 901  | CTGTACTCTTCTGACAGGCATCTGGGGCGGCAATCCAGGGCAATTGTGTGCTTCTCCTCAGG     | 960  |
| Db | 901  | CTGTACTCTTCTGACAGGCATCTGGGGCGGCAATCCAGGGCAATTGTGTGCTTCTCCTCAGG     | 960  |
| Qy | 961  | AGTACCAAGAGCCGCTGTATCCCAAGTGAGGTCAAGCTTGTGGAGACGCTCGGCTCATGTGTC    | 1020 |
| Db | 961  | AGTACCAAGAGCCGCTGTATCCCAAGTGAGGTCAAGCTTGTGGAGACGCTCGGCTCATGTGTC    | 1020 |
| Qy | 1021 | TTTGAACAAGATCGAAGGGAGCTGTGGCGTCTGTGTCTCTTGCGCTCCAAAGCCAGGGTA       | 1080 |
| Db | 1021 | TTTGAACAAGATCGAAGGGAGCTGTGGCGTCTGTGTCTCTTGCGCTCCAAAGCCAGGGTA       | 1080 |
| Qy | 1081 | GCGGAGCTCAAGCTTGGAGAGATATGGGGCTCTCAAGGGCACTGACCCATCTCGAGGTGGAC     | 1140 |
| Db | 1081 | GCGGAGCTCAAGCTTGGAGAGATATGGGGCTCTCAAGGGCACTGACCCATCTCGAGGTGGAC     | 1140 |
| Qy | 1141 | GTGGGAAACGAGCGGCGCCCAATGTGACAGTCTGGGCTTCTTCTGTGTGTGAACGAGAGGAGCTG  | 1200 |
| Db | 1141 | GTGGGAAACGAGCGGCGCCCAATGTGACAGTCTGGGCTTCTTCTGTGTGTGAACGAGAGGAGCTG  | 1200 |
| Qy | 1201 | CCCCACACCCAGAGGCTGTGGAGATCATCTCCGTGTGTGATTTGCCACAGAGCGGTTTC        | 1260 |
| Db | 1201 | CCCCACACCCAGAGGCTGTGGAGATCATCTCCGTGTGTGATTTGCCACAGAGCGGTTTC        | 1260 |
| Qy | 1261 | TTGGCCCGCAATCTGTGCCAAGACTGTGTGGCGGCAAGAGCTGCCCGTGTGAACCGGATGTGGGA  | 1320 |
| Db | 1261 | TTGGCCCGCAATCTGTGCCAAGACTGTGTGGCGGCAAGAGCTGCCCGTGTGAACCGGATGTGGGA  | 1320 |
| Qy | 1321 | GGCGGGAGACAAGACTTGTGGCCGATGTGGCCGATGTGACACTTGTGCTATGATGTGAGACA     | 1380 |
| Db | 1321 | GGCGGGAGACAAGACTTGTGGCCGATGTGGCCGATGTGACACTTGTGCTATGATGTGAGACA     | 1380 |
| Qy | 1381 | CACCTTGTGTGGGGATCTCTGTCTCTCGGGGAGCTGTGTGTGACAGCCGCCCACTGTCTTC      | 1440 |
| Db | 1381 | CACCTTGTGTGGGGATCTCTGTCTCTCGGGGAGCTGTGTGTGACAGCCGCCCACTGTCTTC      | 1440 |

1441 CCGAGCGGAAACCGGGTCTGTCCGATGAGGAGTGTGGCGGTCGGTGGCCAGGCC 1500  
1441 CCGAGCGGAAACCGGGTCTGTCCGATGAGGAGTGTGGCGGTCGGTGGCCAGGCC 1500  
1501 TCTCCCAACGGTCTGACAGTGGGGTGCAGGCTGTGATCTACACAGGGGGCTATCTTCC 1560  
1501 TCTCCCAACGGTCTGACAGTGGGGTGCAGGCTGTGATCTACACAGGGGGCTATCTTCC 1560  
1561 TTTGGGGACCCCAACAGGAGAGAAACGCAACGATATGGCTGTGATCTACCTCTCCAGT 1620  
1561 TTTGGGGACCCCAACAGGAGAGAAACGCAACGATATGGCTGTGATCTACCTCTCCAGT 1620  
1621 CCCCTGCCCTCAAGAAATATACAGCTGTGTGTCCTCCAGCTGGCGGCAAGCCCTG 1680  
1621 CCCCTGCCCTCAAGAAATATACAGCTGTGTGTCCTCCAGCTGGCGGCAAGCCCTG 1680  
1681 GTGATGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCACTATAGGCCAAG 1740  
1681 GTGATGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCACTATAGGCCAAG 1740  
1741 GCGGGGCTACTCCAGAGGCTGAGTCCCAATATACAGATGATGTCTGCAATGGGCT 1800  
1741 GCGGGGCTACTCCAGAGGCTGAGTCCCAATATACAGATGATGTCTGCAATGGGCT 1800  
1801 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCGAGGGTGC 1860  
1801 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCGAGGGTGC 1860  
1861 ATTATGCTGTCCAGGGGCAACAGGGGTGTCTCTTGTGTGAGGACAGCATCTCTGG 1920  
1861 ATTATGCTGTCCAGGGGCAACAGGGGTGTCTCTTGTGTGAGGACAGCATCTCTGG 1920  
1921 ACGCCAGCTTGGCGGCTGTGTGGCATTTGTGATTGGGGCACTGGCTGTGCTGGCCAG 1980  
1921 ACGCCAGCTTGGCGGCTGTGTGGCATTTGTGATTGGGGCACTGGCTGTGCTGGCCAG 1980  
1981 AAGCAGGCGCTCTACACCAAGATGATGATCTTCGAGGATGATCTTCAGAGCCATAAG 2040  
1981 AAGCAGGCGCTCTACACCAAGATGATGATCTTCGAGGATGATCTTCAGAGCCATAAG 2040  
2041 ACTCACTCCGAAGCCAGGAGATGATGATCTTCGAGGATGATCTTCGAGGATGATCTTC 2100  
2041 ACTCACTCCGAAGCCAGGAGATGATGATCTTCGAGGATGATCTTCGAGGATGATCTTC 2100  
2101 GCTTCAGAGGCGCCAGAGTATCCGGTGTGGATCCAGCTGGGCGAGGATGGAGCT 2160  
2101 GCTTCAGAGGCGCCAGAGTATCCGGTGTGGATCCAGCTGGGCGAGGATGGAGCT 2160  
2161 TTTCTCTTGGGCGCGGTCACAGATCCAGATCCAGCTTCAGAGGATCTCTCTTC 2220  
2161 TTTCTCTTGGGCGCGGTCACAGATCCAGATCCAGCTTCAGAGGATCTCTCTTC 2220  
2221 CACAGTGGCGGCGCCACTCAGCCCGAGACCAACCACTCACTCTCTGACCCCATGT 2280  
2221 CACAGTGGCGGCGCCACTCAGCCCGAGACCAACCACTCACTCTCTGACCCCATGT 2280  
2281 AAAATATTGTCTGTCTGTGAGGATCTCTGTCTAGAGTGCCTGTATGATGAGTCTCTTT 2340  
2281 AAAATATTGTCTGTCTGTGAGGATCTCTGTCTAGAGTGCCTGTATGATGAGTCTCTTT 2340  
2341 AAAATATTAAAGATGTTTGAAT 2363  
2341 AAAATATTAAAGATGTTTGAAT 2363

RESULT 3  
US-10-099-322-39/c  
; Sequence 39, Application US/10099322  
; Publication No. US20030215449A1  
; GENERAL INFORMATION:  
; APPLICANT: Mezes et al.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-240CIP  
; CURRENT APPLICATION NUMBER: US/10/099,322  
; CURRENT FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: 60/261,014  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/261,018  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/318,410  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/261,013  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/261,026  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/261,029  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/313,170  
; PRIOR FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: 10/044,564  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 324  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 39  
; LENGTH: 2363  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-099-322-39

Query Match 100.0%; Score 2363; DB 17; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGGGCCACAGTGGAGCGCTGATGC 60  
2363 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGGGCCACAGTGGAGCGCTGATGC 2304  
61 AGGCCCAACGCAACCGCTCTGCTTCAGGCGCGCTGCTGTGGGGCAACATGCTCC 120  
2303 AGGCCCAACGCAACCGCTCTGCTTCAGGCGCGCTGCTGTGGGGCAACATGCTCC 2244  
121 TGCCCAAGCTTGAAGATGACCCGACCCGCGCACTACTGAGAGCTCCGCCCACTGC 180  
2243 TGCCCAAGCTTGAAGATGACCCGACCCGCGCACTACTGAGAGCTCCGCCCACTGC 2184  
181 TGACCCCAAGGATGAGGACCAAGGGGCCCAAGACTCAAGTTCCAGCCCTGAGGACAGGG 240  
2183 TGACCCCAAGGATGAGGACCAAGGGGCCCAAGACTCAAGTTCCAGCCCTGAGGACAGGG 2124  
241 TTCCCTATCCCGCAACCGCTTAATGCCACCTCTTAATGAGGGGTTCTGGGAC 300  
2123 TTCCCTATCCCGCAACCGCTTAATGCCACCTCTTAATGAGGGGTTCTGGGAC 2064  
301 TGAAGAGGGGCACTATGAGAGTCTCCCAAGCACTAGAGTCTCTGCTCTTCTT 360  
2063 TGAAGAGGGGCACTATGAGAGTCTCCCAAGCACTAGAGTCTCTGCTCTTCTT 2004  
361 CAGACTCAGCGGTGAGACCCCAAGTCTTCTCTCCCAAGCACTAGAGTCTCAGCCCTCAG 420  
2003 CAGACTCAGCGGTGAGACCCCAAGTCTTCTCTCCCAAGCACTAGAGTCTCAGCCCTCAG 1944  
421 GCCCTCTCTCTATATGAGGAGTCTGAGCCCGCAAAATTTCTCTTTCCAAAGCTTA 480  
1943 GCCCTCTCTCTATATGAGGAGTCTGAGCCCGCAAAATTTCTCTTTCCAAAGCTTA 1884  
481 TGATTTAAGTCTCAGAGTCTCTCTCTCCCAAGCACTAGAGTCTCTGCTCTCAG 540  
1883 TGATTTAAGTCTCAGAGTCTCTCTCTCCCAAGCACTAGAGTCTCTGCTCTCAG 1824  
541 CAGGCTCAGGATGAGGGGTCCTCCATCTCTGCAATTCAGAGGTCCTCCGCTGCTGCA 600  
1823 CAGGCTCAGGATGAGGGGTCCTCCATCTCTGCAATTCAGAGGTCCTCCGCTGCTGCA 1764  
601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTGCTGATCAAGGCTGCTTGG 660

|    |      |   |      |
|----|------|---|------|
| Db | 1763 | GACACGTACCCCATCTCTTGAACCCAGGCCCAATCTGGGTCCGATACAGAGCGTGTCTCG    | 1704 |
| Qy | 661  | CCAAAGCCCACTGTCCTTAACAGCTGCTGGATGAGACGCTTGAGACTGGGAGCGCCAGAGACT | 720  |
| Db | 1703 | CCAAAGCCCACTGTCCTTAACAGCTGCTGGATGAGACGCTTGAGACTGGGAGCGCCAGAGACT | 1644 |
| Qy | 721  | GGGCTGGGCTGGGGCTCCCCCAGGCCCTGCTCCGCCCTCACTTCTTCAAGGTCCCAACC     | 780  |
| Db | 1643 | GGGCTGGGCTGGGGCTCCCCCAGGCCCTGCTCCGCCCTCACTTCTTCAAGGTCCCAACC     | 1584 |
| Qy | 781  | TGGCCAGAGAGGTCAAGCCAGGAGATCATTTAAACAAGAGCAGTGAATGAGCGCAGAAAGAG  | 840  |
| Db | 1563 | TGGCCAGAGAGGTCAAGCCAGGAGATCATTTAAACAAGAGCAGTGAATGAGCGCAGAAAGAG  | 1524 |
| Qy | 841  | GGTGGCCGAGACTGTGCTCAGTCTCCAGACCCAAAGTGGACACTTCACTGGCGGAGAC      | 900  |
| Db | 1523 | GGTGGCCGAGACTGTGCTCAGTCTCCAGACCCAAAGTGGACACTTCACTGGCGGAGAC      | 1464 |
| Qy | 901  | CTGCTAATTTCTTGAACAGCCATCGGGGGGCAATCTGGGGCCATTGTGGCTGTCTCTCAAG   | 960  |
| Db | 1463 | CTGCTAATTTCTTGAACAGCCATCGGGGGGCAATCTGGGGCCATTGTGGCTGTCTCTCAAG   | 1404 |
| Qy | 961  | AGTGAACAGAGACCCGCTGTACCCAGATGACAGGTCAAGCTCTGCGGAGACGCTCGAGTCA   | 1020 |
| Db | 1403 | AGTGAACAGAGACCCGCTGTACCCAGATGACAGGTCAAGCTCTGCGGAGACGCTCGAGTCA   | 1344 |
| Qy | 1021 | TTTGAACAAGACGGAAGGACGTGGCGGCTGTGCTCTCGCGCTCCAAAGCCAGGTA         | 1088 |
| Db | 1343 | TTTGAACAAGACGGAAGGACGTGGCGGCTGTGCTCTCGCGCTCCAAAGCCAGGTA         | 1284 |
| Qy | 1081 | GCCGGAATCAGCTGGGAGAGATGGGGCTTCTCAAGGGAATGACCCATCCGAGCTGGAC      | 1140 |
| Db | 1283 | GCCGGAATCAGCTGGGAGAGATGGGGCTTCTCAAGGGAATGACCCATCCGAGCTGGAC      | 1224 |
| Qy | 1141 | GTGCAAAAGGCGGCGGCGCCAAATGGACGTCGAGGCTTCTTCTGTGTGGAAGAAGGAGGCTG  | 1200 |
| Db | 1223 | GTGCAAAAGGCGGCGGCGCCAAATGGACGTCGAGGCTTCTTCTGTGTGGAAGAAGGAGGCTG  | 1164 |
| Qy | 1201 | CCCCAACCCAGAGGCTGTGAGGTCAATCTCCGTGTGATTTGGCCCAAGAGCGCTTTC       | 1260 |
| Db | 1163 | CCCCAACCCAGAGGCTGTGAGGTCAATCTCCGTGTGATTTGGCCCAAGAGCGCTTTC       | 1104 |
| Qy | 1261 | TTTGGCCGCAATCTGCAAGAATGTGTGGCCGCAAGAAAGCTGCCCGTGGACCGCATGTGGGA  | 1320 |
| Db | 1103 | TTTGGCCGCAATCTGCAAGAATGTGTGGCCGCAAGAAAGCTGCCCGTGGACCGCATGTGGGA  | 1044 |
| Qy | 1321 | GGCGGGGACACCACTTTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTCCGTAATGATGAGACA  | 1380 |
| Db | 1043 | GGCGGGGACACCACTTTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTCCGTAATGATGAGACA  | 984  |
| Qy | 1381 | CACCTCTGTGGGGAGATCCCTGCTCTCCGGGAGCTGGAGCTGACAGCGCCCACTGCTTC     | 1440 |
| Db | 983  | CACCTCTGTGGGGAGATCCCTGCTCTCCGGGAGCTGGAGCTGACAGCGCCCACTGCTTC     | 924  |
| Qy | 1441 | CCGAGCGGAAACCGGGTCTCTGCCGATGGCGAGTGTGGCGGCTGGGCCCAAGCC          | 1500 |
| Db | 923  | CCGAGCGGAAACCGGGTCTCTGCCGATGGCGAGTGTGGCGGCTGGGCCCAAGCC          | 864  |
| Qy | 1501 | TCTCCCAACGGTCTGCAAGCTGGAGGTGAGAGCTGTGGTCAACACAGGGGGCTAATCTTCCC  | 1560 |
| Db | 863  | TCTCCCAACGGTCTGCAAGCTGGAGGTGAGAGCTGTGGTCAACACAGGGGGCTAATCTTCCC  | 804  |
| Qy | 1561 | TTTGGGAGACCCCAACAGCGAGAGAGAACAGCAACGATTTATGGCTTGGTCACTCTCCAGT   | 1620 |
| Db | 803  | TTTGGGAGACCCCAACAGCGAGAGAGAACAGCAACGATTTATGGCTTGGTCACTCTCCAGT   | 744  |
| Qy | 1621 | CCCTGCCCCCTACAGAAATACATCCAGCTGTGTGCTTCCAGCTGCCGCGCAGAGCCCTG     | 1680 |
| Db | 743  | CCCTGCCCCCTACAGAAATACATCCAGCTGTGTGCTTCCAGCTGCCGCGCAGAGCCCTG     | 684  |
| Qy | 1681 | GTGATATGCAAGATTTGTAACGTGACGGGCTGGGGCAACAGCAGTACTAATGCGCAACAG    | 1740 |
| Db | 683  | GTGATATGCAAGATTTGTAACGTGACGGGCTGGGGCAACAGCAGTACTAATGCGCAACAG    | 624  |

|  |      |      |             |          |            |             |               |         |          |                      |       |      |
|--|------|------|-------------|----------|------------|-------------|---------------|---------|----------|----------------------|-------|------|
| QY   | 1741 | GC   | CGGCGGAGTAC | CC       | CAGAGAGGCT | CGAGTCC     | CCCATTAAT     | CAGCANT | AGTCTG   | CAATAGGCGCT          | 1800  |      |
| Db   | 623  | GC   | CGGCGGAGTAC | CC       | CAGAGAGGCT | CGAGTCC     | CCCATTAAT     | CAGCANT | AGTCTG   | CAATAGGCGCT          | 564   |      |
| QY   | 1801 | GACT | CTCTATG     | GA       | AAAC       | CCAGATCAAG  | CCCAAGATGTTCT | TGCTG   | CTG      | CTACCCCGAGAGTGGC     | 1868  |      |
| Db   | 563  | GACT | CTCTATG     | GA       | AAAC       | CCAGATCAAG  | CCCAAGATGTTCT | TGCTG   | CTG      | CTACCCCGAGAGTGGC     | 504   |      |
| QY   | 1861 | ATTG | TGCTCTGC    | CA       | AGGCGG     | CAAGGCGG    | TGTCCTTT      | TGTGTG  | AGAGAC   | AGATCTCTGG           | 1920  |      |
| Db   | 503  | ATTG | TGCTCTGC    | CA       | AGGCGG     | CAAGGCGG    | TGTCCTTT      | TGTGTG  | AGAGAC   | AGATCTCTGG           | 444   |      |
| QY   | 1921 | ACG  | CCACGTTGG   | CCGCGCTG | TGGGCA     | TTGTGAGTTGG | GGGCACTTGG    | CTGTGG  | CCCTGG   | CCGAC                | 1980  |      |
| Db   | 443  | ACG  | CCACGTTGG   | CCGCGCTG | TGGGCA     | TTGTGAGTTGG | GGGCACTTGG    | CTGTGG  | CCCTGG   | CCGAC                | 384   |      |
| QY   | 1961 | AAG  | CCAGGCGT    | CTA      | CA         | CCCAAGTCA   | GTGACTTCC     | GGAGTGG | ATCTTCC  | CAGGCCATAAG          | 2040  |      |
| Db   | 363  | AAG  | CCAGGCGT    | CTA      | CA         | CCCAAGTCA   | GTGACTTCC     | GGAGTGG | ATCTTCC  | CAGGCCATAAG          | 324   |      |
| QY   | 2041 | ACT  | CAC         | TCTCC    | GAAGCAG    | CGGCGATG    | GTGACCC       | AGCTCT  | GA       | CCGGTGGCTTCTCGCTGGCA | 2100  |      |
| Db   | 323  | ACT  | CAC         | TCTCC    | GAAGCAG    | CGGCGATG    | GTGACCC       | AGCTCT  | GA       | CCGGTGGCTTCTCGCTGGCA | 264   |      |
| QY   | 2101 | GCT  | TC          | CAGGGG   | CCCGAGG    | TGATCCCG    | GGTGGAGTCC    | ACGCTGG | GGGCGAGG | ATGGAGCGT            | 2160  |      |
| Db   | 263  | GCT  | TC          | CAGGGG   | CCCGAGG    | TGATCCCG    | GGTGGAGTCC    | ACGCTGG | GGGCGAGG | ATGGAGCGT            | 204   |      |
| QY   | 2161 | TTTT | CTTCTT      | TGGG     | CCCGG      | TCC         | CAGAGT        | CCAA    | GGA      | CACCCCTCC            | CTC   | 2220 |
| Db   | 203  | TTTT | CTTCTT      | TGGG     | CCCGG      | TCC         | CAGAGT        | CCAA    | GGA      | CACCCCTCC            | CTC   | 144  |
| QY   | 2221 | CAC  | AGTGG       | CGGG     | CCCA       | CTC         | CAG           | CCCCG   | AGACCA   | CCCA                 | ACTCA | 2280 |
| Db   | 143  | CAC  | AGTGG       | CGGG     | CCCA       | CTC         | CAG           | CCCCG   | AGACCA   | CCCA                 | ACTCA | 84   |
| QY   | 2281 | AAA  | TATTTG      | TCTG     | CTG        | CTG         | TGGG          | CA      | CTC      | GTCT                 | GA    | 2340 |
| Db   | 83   | AAA  | TATTTG      | TCTG     | CTG        | CTG         | TGGG          | CA      | CTC      | GTCT                 | GA    | 24   |
| QY   | 2341 | AAAT | ATATTA      | AGATG    | GT         | TTTG        | ATT           | TG      | ATT      | 2363                 |       |      |
| Db   | 23   | AAAT | ATATTA      | AGATG    | GT         | TTTG        | ATT           | TG      | ATT      | 2363                 |       |      |
| RESULT 4   |      |      |             |          |            |             |               |         |          |                      |       |      |
| US-10-044-564-39/c   |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; Sequence 39, Application US/10044564                         |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; Publication No. US20040018196A1                              |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; GENERAL INFORMATION:   |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; APPLICANT: Mezes et al.                                      |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; FILE REFERENCE: 21402-240                                    |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; CURRENT APPLICATION NUMBER: US/10/044,564                    |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; CURRENT FILING DATE: 2002-09-09                              |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; PRIOR APPLICATION NUMBER: 60/261,014                         |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; PRIOR FILING DATE: 2001-01-11                                |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; PRIOR APPLICATION NUMBER: 60/261,018                         |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; PRIOR FILING DATE: 2001-01-11                                |      |      |             |          |            |             |               |         |          |                      |       |      |
| ; PRIOR APPLICATION NUMBER: 60/318,410                         |      |      |             |          |            |             |               |         |          |                      |       |      |

LENGTH: 2363  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-044-564-39

Query Match 100.0%; Score 2363; DB 17; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAGCCCGCTTTCCAGGAGACCTTACCTGAGGGGCCCAAGGTGAGGGCAGCTGGGCTTACG 60  
DB 2363 TCGAGCCCGCTTTCCAGGAGACCTTACCTGAGGGGCCCAAGGTGAGGGCAGCTGGGCTTACG 2304  
OY 61 AGGCCCCAGCCACGCGCTTGCCTCCAGGCGCGCGCTGCTGCGGGGCCAATGCTCC 120  
DB 2303 AGGCCCCAGCCACGCGCTTGCCTCCAGGCGCGCGCTGCTGCGGGGCCAATGCTCC 2244  
OY 121 TCGCCAGGCTTGAGAGCTGACCCGACCCCGGCACTACTGAGGCTCCGCCCACTGC 180  
DB 2243 TCGCCAGGCTTGAGAGCTGACCCGACCCCGGCACTACTGAGGCTCCGCCCACTGC 2184  
OY 181 TCGAGCCCGAGGTAAAGGACAAAGGGCCCCAGACTCAAGTTCCAGCCTGAGAGAGGGG 240  
DB 2183 TCGAGCCCGAGGTAAAGGACAAAGGGCCCCAGACTCAAGTTCCAGCCTGAGAGAGGGG 2124  
OY 241 TTCCCTCATCCCCCAGCCAGCTTAATGCCCACCTCTAATGAGGGGTTCTGGGAGC 300  
DB 2123 TTCCCTCATCCCCCAGCCAGCTTAATGCCCACCTCTAATGAGGGGTTCTGGGAGC 2064  
OY 301 TGAAGAGGGGCACTATGACGTCTCCCAAGACCTAGAGTGTCTGTCTGTCTTCTT 360  
DB 2063 TGAAGAGGGGCACTATGACGTCTCCCAAGACCTAGAGTGTCTGTCTGTCTTCTT 2004  
OY 361 CAGACTGAGCGTTGGAACCCAGTCTTTCTCCCAAGACCCAGAGATTCCAGCCTCAG 420  
DB 2003 CAGACTGAGCGTTGGAACCCAGTCTTTCTCCCAAGACCCAGAGATTCCAGCCTCAG 1944  
OY 421 GCCCCTCTCCCTCATCTAGAGGAGTCCGAGCCCCCAATTCTCTTCCCAAGACTTA 480  
DB 1943 GCCCCTCTCCCTCATCTAGAGGAGTCCGAGCCCCCAATTCTCTTCCCAAGACTTA 1884  
OY 481 TGATTTCAGGTCTTCAAGTGTCTCTCCCTCAAAACGGGATCTCTAGTCCCTGTCTCAG 540  
DB 1883 TGATTTCAGGTCTTCAAGTGTCTCTCCCTCAAAACGGGATCTCTAGTCCCTGTCTCAG 1824  
OY 541 CAGGCTCAGGCATGGGGGTCCCCATCTCTGAAATCCAGGGTCCCCCGCTGCTGCA 600  
DB 1823 CAGGCTCAGGCATGGGGGTCCCCATCTCTGAAATCCAGGGTCCCCCGCTGCTGCA 1764  
OY 601 GACACTGAGCCCATCTTGAACCCAGCCCAATCTGCTCCGTATCAAGGCGTGTCTGG 660  
DB 1763 GACACTGAGCCCATCTTGAACCCAGCCCAATCTGCTCCGTATCAAGGCGTGTCTGG 1704  
OY 661 CCAGAGCCCAATCTCTTACAGCCTGCTGATGAGCGCTGAGACTGGGGGCGCCAGACT 720  
DB 1703 CCAGAGCCCAATCTCTTACAGCCTGCTGATGAGCGCTGAGACTGGGGGCGCCAGACT 1644  
OY 721 GGGCTGGGCTGGGCTCCCCAGGCGCTTCCCTGCTCATCTTCTCAAGGTCCCAACC 780  
DB 1643 GGGCTGGGCTGGGCTCCCCAGGCGCTTCCCTGCTCATCTTCTCAAGGTCCCAACC 1584  
OY 781 TGGCCAGAGAGGTGAGCCAGGGGAATCATTAACAAGGAGGATGACATGGCGAGAAAGAG 840  
DB 1583 TGGCCAGAGAGGTGAGCCAGGGGAATCATTAACAAGGAGGATGACATGGCGAGAAAGAG 1524  
OY 841 GGTGGCCGGAAGCTGTGCACTGTCTCCAGACCCAGAGGTGAGACTCTCACTGCGGGAGC 900  
DB 1523 GGTGGCCGGAAGCTGTGCACTGTCTCCAGACCCAGAGGTGAGACTCTCACTGCGGGAGC 1464  
OY 901 CTGCTACTTCTGACAGGCACTGGGGGGGCAATCTGGGCAATTTGGCTGTCTCTCTCAGG 960  
DB 1463 CTGCTACTTCTGACAGGCACTGGGGGGGCAATCTGGGCAATTTGGCTGTCTCTCTCAGG 1404

OY 961 AGTACAGAGAGCGCTGTATCCAGTGCAGGTCACTCTGCGAGCGCTCGGCTCATGTGCT 1020  
DB 1403 AGTACAGAGAGCGCGTGTATCCAGTGCAGGTCACTCTGCGAGCGCTCGGCTCATGTGCT 1344  
OY 1021 TTTGACAAAGCGAAGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAAAGCCAGGTTA 1080  
DB 1343 TTTGACAAAGCGAAGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAAAGCCAGGTTA 1284  
OY 1081 GCGGAGCTCAGCTGCGAGGAGATGGGGCTTCTCGAGGGCACTGACCCACTCCGAGTGGAC 1140  
DB 1283 GCGGAGCTCAGCTGCGAGGAGATGGGGCTTCTCGAGGGCACTGACCCACTCCGAGTGGAC 1224  
OY 1141 GTGGAACGCGCGGCGCCAAATGACAGTCCGAGCTTCTTCTGTGTGAAGAGGAGCTG 1200  
DB 1223 GTGGAACGCGCGGCGCCAAATGACAGTCCGAGCTTCTTCTGTGTGAAGAGGAGGCTG 1164  
OY 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCCGTGTGTATTTGCCCAAGAGGCGTTTC 1260  
DB 1163 CCCCACACCCAGAGGCTGTGAGAGTATCTCCGTGTGTATTTGCCCAAGAGGCGGTTTC 1104  
OY 1261 TTGGCGGCAATCTGSCAAAGACTGTGGCCGACAGAGAGGTCGCCGTGACCGGATGTGGGA 1320  
DB 1103 TTGGCGGCAATCTGSCAAAGACTGTGGCCGACAGAGAGGTCGCCGTGACCGGATGTGGGA 1044  
OY 1321 GCGCGGAGCACCAAGCTTGGGCGGCTGGCCGTGCAAGTCAAGCTTGTCTATGATGAGACA 1380  
DB 1043 GCGCGGAGCACCAAGCTTGGGCGGCTGGCCGTGCAAGTCAAGCTTGTCTATGATGAGACA 984  
OY 1381 CACTCTGTGGGGGATCTCTGTCTCTCGGGGACCTGGGTGTGACAGCGCGCCACTGCTTC 1440  
DB 983 CACTCTGTGGGGGATCTCTGTCTCTCGGGGACCTGGGTGTGACAGCGCGCCACTGCTTC 924  
OY 1441 CCGGAGGGGAACCGGGGCTGTGCCGATGGAGATGTTTGGCGGTGGCCGAGGCC 1500  
DB 923 CCGGAGGGGAACCGGGGCTGTGCCGATGGAGATGTTTGGCGGTGGCCGAGGCC 864  
OY 1501 TCTCCCAAGGTCTGACAGCTGGGGGTCAAGGCTGTGTCTTACAAGGGGCTATCTTCCC 1560  
DB 863 TCTCCCAAGGTCTGACAGCTGGGGGTCAAGGCTGTGTGTCTTACAAGGGGCTATCTTCCC 804  
OY 1561 TTTTGGGACCCCAACAGGAGAGAGACAGACATTTTGGCTGTGTCTTCACTTCCAGT 1620  
DB 803 TTTTGGGACCCCAACAGGAGAGAGACAGACATTTTGGCTGTGTCTTCACTTCCAGT 744  
OY 1621 CCGCGGCCCTCAAGAAATCATCCAGCTGTGTGTCGCCAGCTGCGGCGCAGGCCCTG 1680  
DB 743 CCGCGGCCCTCAAGAAATCATCCAGCTGTGTGTCGCCAGCTGCGGCGCAGGCCCTG 684  
OY 1681 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGGCCAAG 1740  
DB 683 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGGCCAAG 624  
OY 1741 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATATTTGTGTGAAATGGCGCT 1800  
DB 623 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATATTTGTGTGAAATGGCGCT 564  
OY 1801 GACTTCTATGGAACCCAGATCAAGCCCAAGATGTTCTGTGTGCTGACTCCCGAGGGTGGC 1860  
DB 563 GACTTCTATGGAACCCAGATCAAGCCCAAGATGTTCTGTGTGCTGACTCCCGAGGGTGGC 504  
OY 1861 ATTGATCTGTGCGAGGCGACAGGCGTGTCTCTTGTGTGTGAGGACAGCATCTTCGG 1920  
DB 503 ATTGATCTGTGCGAGGCGACAGGCGTGTCTCTTGTGTGTGAGGACAGCATCTTCGG 444  
OY 1921 ACGCCAGTGTGGCGCTGTGTGCAATTTGTGATTTGGGCACTGGCTGTGCTTGGCCAG 1980  
DB 443 ACGCCAGTGTGGCGCTGTGTGCAATTTGTGATTTGGGCACTGGCTGTGCTTGGCCAG 384  
OY 1981 AAGCAGGCGCTTACCAAAAGTGAAGCTTCCGGAGTGGAGCTTCCAGGCGCATTAAG 2040  
DB 383 AAGCAGGCGCTTACCAAAAGTGAAGCTTCCGGAGTGGAGCTTCCAGGCGCATTAAG 324  
OY 2041 ACTCATCTCCGAAGCCAGCGGATGTGACCAAGCTTGACCGGTGTCTTCTGCTGCGCA 2100







1216 CTGCTGAGAGATCATCTCCGTGTGTGATTGCCCCAGAGAGCCGTTTCTTGCCCGCCATCTGC 1275  
 636 CTGCTGAGAGATCATCTCCGTGTGTGATTGCCCCAGAGAGCCGTTTCTTGCCCGCCATCTGC 695  
 1276 CAAGACTGTGCGCCGAGAGAGTGCCTGTGACCCGATCTGTGAGAGCCGGAGACACAGC 1335  
 696 CAAGACTGTGCGCCGAGAGAGTGCCTGTGACCCGATCTGTGAGAGCCGGAGACACAGC 755  
 1336 TTGGGCGCGGTGCGCTGTGACAGCTTGTGATGATGAGACACACTCTGTGTGGGGA 1395  
 756 TTGGGCGCGGTGCGCTGTGACAGCTTGTGATGATGAGACACACTCTGTGTGGGGA 815  
 1396 TCCCTGCTCTCCGGGAGCTGGGTGTGACAGCCGCGCATCTGTTCCCGGAGCGGAACCG 1455  
 816 TCCCTGCTCTCCGGGAGCTGGGTGTGACAGCCGCGCATCTGTTCCCGGAGCGGAACCG 875  
 1456 GTCTGTCTCCGATGTGCGAGTGTGTCCTGGATGCGGTGCGGAGGCTCTCCACAGGTGTG 1515  
 876 GTCTGTCTCCGATGTGCGAGTGTGTCCTGGATGCGGTGCGGAGGCTCTCCACAGGTGTG 935  
 1516 CAGCTGGGGGTGACAGCTGTGTGCTACACAGGGGGCTATCTTCCCTTGTGGGAGCCCAAC 1575  
 936 CAGCTGGGGGTGACAGCTGTGTGCTACACAGGGGGCTATCTTCCCTTGTGGGAGCCCAAC 995  
 1576 AGCGAGGAGAACAGAACGATATTGGCTGTGCTCCATCTCTCCAGTCCCTCTCCATCA 1635  
 996 AGCGAGGAGAACAGAACGATATTGGCTGTGCTCCATCTCTCCAGTCCCTCTCCATCA 1055  
 1636 GAATACATCCAGCTGTGTGCTCCATCTCTCCAGTCCCGGAGGCTGTGTGATGAGCAAGATC 1695  
 1056 GAATACATCCAGCTGTGTGCTCCATCTCTCCAGTCCCGGAGGCTGTGTGATGAGCAAGATC 1115  
 1696 TGTACCTGTGACGGGTGTGGGAGAACAGAGTATATGAGCAACAGCCGGGGGTACTCCAG 1755  
 1116 TGTACCTGTGACGGGTGTGGGAGAACAGAGTATATGAGCAACAGCCGGGGGTACTCCAG 1175  
 1756 GAGGCTGAGTCCCATATATACAGCAATGATCTGCAATGAGCGCTGATCTTATAGGAAC 1815  
 1176 GAGGCTGAGTCCCATATATACAGCAATGATCTGCAATGAGCGCTGATCTTATAGGAAC 1235  
 1816 CAGATCAAGCCCAAGATGTTCTGTGTGCTGCTACCCCGAGGGGTGAGTATGATCCAGC 1875  
 1236 CAGATCAAGCCCAAGATGTTCTGTGTGCTGCTACCCCGAGGGGTGAGTATGATGCTGCGAG 1295  
 1876 GCGCAGACGGGTGTGCTGTTGTGTGTGAGAACAGCATCTCTCGAGCGCCAGCTTGGCGG 1935  
 1296 GCGCAGACGGGTGTGCTGTTGTGTGTGAGAACAGCATCTCTCGAGCGCCAGCTTGGCGG 1355  
 1936 CTGTGTGCGATTTGTGAGTTGGGGGCACTGGCTGTGCTGCGCCAGAGCCAGGGGTCTAC 1995  
 1356 CTGTGTGCGATTTGTGAGTTGGGGGCACTGGCTGTGCTGCGCCAGAGCCAGGGGTCTAC 1415  
 1996 ACCAAGTCAGTGACTTCGGGAGGTGATCTTCCAGGGCATTAAGACATCTCCGAAGCC 2055  
 1416 ACCAAGTCAGTGACTTCGGGAGGTGATCTTCCAGGGCATTAAGACATCTCCGAAGCC 1475  
 2056 AGCGGATGTGATCCAGCTGTGACCGGTGTGCTCTGCTGCGAGCCCTCCAGGGGCCGA 2115  
 1476 AGCGGATGTGATCCAGCTGTGACCGGTGTGCTCTGCTGCGAGCCCTCCAGGGGCCGA 1535  
 2116 GGTGATCTCCGGGTGTGGGATTCACGCTGTGGGCGGAGATGGGACGTTTTTCTTGTGGGCC 2175  
 1536 GGTGATCTCCGGGTGTGGGATTCACGCTGTGGGCGGAGATGGGACGTTTTTCTTGTGGGCC 1595  
 2176 CGGTTCACAGGTTCAGAGACACCTCTCCAGGGGTCTCTCTTCCACAGTGTGGGGGCC 2235  
 1596 CGGTTCACAGGTTCAGAGACACCTCTCCAGGGGTCTCTCTTCCACAGTGTGGGGGCC 1655  
 2236 ACTCAGCCCGAGACACCACTGTGACCTCTGTGACCCCGCATGTAAATATTTGTCT 2295  
 1656 ACTCAGCCCGAGACACCACTGTGACCTCTGTGACCCCGCATGTAAATATTTGTCT 1715  
 2296 GTCTGGGAGCTCTGTGTAGTGTCCCTGTATGATGGGATGCTCTTAAATATAAGATGG 2355

1716 GTCTGAGAGATCATCTCCGTGTGTGATTGCCCCAGAGAGCCGTTTCTTGCCCGCCATCTGC 1775  
 2356 TTTTGATT 2363  
 1776 TTTTGATT 1783

RESULT 7  
 US-09-776-191-67  
 ; Sequence 67, Application US/09776191  
 ; Publication No. US20030119168A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwin L. Madison  
 ; APPLICANT: Edgar O. Ong  
 ; APPLICANT: Jihun-Chern Yeh  
 ; APPLICANT: Corvas International, Inc.  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
 ; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND  
 ; TITLE OF INVENTION: METHODS BASED THEREON  
 ; FILE REFERENCE: 24745-1607  
 ; CURRENT APPLICATION NUMBER: US/09/776.191  
 ; PRIOR APPLICATION NUMBER: 60/213,124  
 ; PRIOR FILING DATE: 2000-06-22  
 ; PRIOR APPLICATION NUMBER: 60/234,840  
 ; PRIOR FILING DATE: 2000-06-22  
 ; PRIOR APPLICATION NUMBER: 60/179,982  
 ; PRIOR FILING DATE: 2000-02-03  
 ; PRIOR APPLICATION NUMBER: 60/183,542  
 ; PRIOR FILING DATE: 2000-02-18  
 ; PRIOR APPLICATION NUMBER: 09/657,968  
 ; PRIOR FILING DATE: 2000-02-08  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 67  
 ; LENGTH: 1783  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (246)..(1499)  
 ; OTHER INFORMATION: Nucleotide sequence encoding human hepsin  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: GenBank M18930  
 ; DATABASE ENTRY DATE: 1993-06-11  
 ; US-09-776-191-67

Query Match 67.4%; Score 1593.6; DB 10; Length 1783;  
 Best Local Similarity 99.1%; Pred. No. 0;  
 Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

736 CCCCCAGGCGCTGCTCCCGGTGCATCTCCACAGGTCCACCTTGGCCAGGAGGTCA 795  
 157 CTTGAGAGCTTCGCCCCCATCTGTGAGACCCA-GGATCCACCTTGGCCAGGAGGTCA 215  
 796 GCCAGGGAATCATTAACAAGAGGAGTGAATGAGCGCAGAGAGAGGGTGTG 855  
 216 GCCAGGGAATCATTAACAAGAGGAGTGAATGAGCGCAGAGAGAGGGTGTG 275  
 856 CCATGCTGCTCCAGACCCCAAGGTGTGAGCTCTCACTGCGGGGACCTCTACTTCTACA 915  
 276 CCAATGCTGCTCCAGACCCCAAGGTGTGAGCTCTCACTGCGGGGACCTCTACTTCTACA 335  
 916 GCCATCGGGGCGGATCTGTGGGCAATTTGGCTGTCTCTCCAGAGTGAACAGAGCCG 975  
 336 GCCATCGGGGCGGATCTGTGGGCAATTTGGCTGTCTCTCCAGAGTGAACAGAGCCG 395  
 976 CTGTACCACTGTGACAGTCTGTGCGAGCGCTGTGCTATGATCTTTTGAACAAGCGAA 1035  
 396 CTGTACCACTGTGACAGTCTGTGCGAGCGCTGTGCTATGATCTTTTGAACAAGCGAA 455  
 1036 GGAGACGTGGGGGCTGCTGTGCTCTGTGCGGCTCCACCGCAGGGATAGCGGAGTCACTGC 1095



1156 GCCAATGGACAGTGGGCTTCTTCTGTGTGACGAGGGGAGGCTGCCCAACCCAGAG 1215  
576 GCCAATGGACAGTGGGCTTCTTCTGTGTGACGAGGGGAGGCTGCCCAACCCAGAG 635  
1216 CTGCTGAGAGTCACTCTCCGTGTGTGATTGGCCAGAGGCGCTTTCTTGGCCGACATGCG 1275  
636 CTGCTGAGAGTCACTCTCCGTGTGTGATTGGCCAGAGGCGCTTTCTTGGCCGACATGCG 695  
1276 CAAGACTGTGGCCGAGGAGAGCTGCGGTGACCGCATCTGTGGAGGCGGGAGACACAGC 1335  
696 CAAGACTGTGGCCGAGGAGAGCTGCGGTGACCGCATCTGTGGAGGCGGGAGACACAGC 755  
1336 TTGGCCCGGTGGCCCTGTGGCAAGTCAAGCTTGTGTGATGTGAGAGCACTCTCTGTGGGGA 1395  
756 TTGGCCCGGTGGCCCTGTGGCAAGTCAAGCTTGTGTGATGTGAGAGCACTCTCTGTGGGGA 815  
1396 TCCCTGCTCTCCGGGGAGCTGGGTGCTGACAGCGCGCCCACTGCTTCCGGAGCGGAAACGG 1455  
816 TCCCTGCTCTCCGGGGAGCTGGGTGCTGACAGCGCGCCCACTGCTTCCGGAGCGGAAACGG 875  
1456 GTCTCTGTCGGAATGGAGAGTGTGTCGCGGTGCGGTGCGGAGGCGCTCTCCGACGCTCTG 1515  
876 GTCTCTGTCGGAATGGAGAGTGTGTCGCGGTGCGGTGCGGAGGCGCTCTCCGACGCTCTG 935  
1516 CAGCTGGGGGTGACAGGCTGTGTCTACACAGGGGCTATCTTCTTCCGGAACCCCAAC 1575  
936 CAGCTGGGGGTGACAGGCTGTGTCTACACAGGGGCTATCTTCTTCCGGAACCCCAAC 995  
1576 AGCGAGAGGAAACAGCAAGATATGCTCCGTGCTCACTCTCAAGTCCCTGCGCCCTCA 1635  
996 AGCGAGAGGAAACAGCAAGATATGCTCCGTGCTCACTCTCAAGTCCCTGCGCCCTCA 1055  
1636 GAATACATCAAGCCTGTGTGCTCCAGCTGCGGCGAGGCGCTGTGTGTGATGCAAGATC 1695  
1056 GAATACATCAAGCCTGTGTGCTCCAGCTGCGGCGAGGCGCTGTGTGTGATGCAAGATC 1115  
1696 TGTACCGGTGACGGGCTGGGGGCAACGCAATATGTGCAACAGGCGGGGTCTCCAG 1175  
1116 TGTACCGGTGACGGGCTGGGGGCAACGCAATATGTGCAACAGGCGGGGTCTCCAG 1175  
1756 GAGGCTGAGTCCCAATATGAGCAATGATGTGTGCAATGTGGGCTGATCTTCTATGAGAAC 1815  
1176 GAGGCTGAGTCCCAATATGAGCAATGATGTGTGCAATGTGGGCTGATCTTCTATGAGAAC 1235  
1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGCTGCCAG 1875  
1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGCTGCCAG 1235  
1876 GGGCAGACGGGTGCTCTTGTGTGTGAGACAGCATCTTCCGACGCGACGTTGGCGG 1935  
1296 GGGCAGACGGGTGCTCTTGTGTGTGAGACAGCATCTTCCGACGCGACGTTGGCGG 1355  
1936 CTGTGTGAGCATTTGTGAGTTGGGGGCACTGGCTGTGCTGGGCGGAGGAGGAGGCTAC 1995  
1356 CTGTGTGAGCATTTGTGAGTTGGGGGCACTGGCTGTGCTGGGCGGAGGAGGAGGCTAC 1415  
1996 ACCAAGTCAAGTCACTTCCGGGAGTGAATCTTCCAGGCGCATTAAGTCACTCCGAAGC 2055  
1416 ACCAAGTCAAGTCACTTCCGGGAGTGAATCTTCCAGGCGCATTAAGTCACTCCGAAGC 1475  
2056 AGCGGATGTGACCCAGCTCTGACCGGTGCTTCCGTGCGACGCTCCAGGGCGGGA 2115  
1476 AGCGGATGTGACCCAGCTCTGACCGGTGCTTCCGTGCGACGCTCCAGGGCGGGA 1535  
2116 GGTGATCCCGGTGGGTGGATCCAGCTGGGCGGAGATGGAGCGTTTCTTCTTGGGCG 2175  
1536 GGTGATCCCGGTGGGTGGATCCAGCTGGGCGGAGATGGAGCGTTTCTTCTTGGGCG 1555  
2176 CGGTTCACAGGTCCAGAGACACCTCCCTCAAGGGTCTCTCTTCCACAGTGGCGGCGC 2235  
1596 CGGTTCACAGGTCCAGAGACACCTCCCTCAAGGGTCTCTCTTCCACAGTGGCGGCGC 1655

2236 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACGCCCCCATATATATTTCTGCT 2295  
1556 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACGCCCCCATATATATTTCTGCT 1715  
2296 GTCTGGAGATCTCTGTCTAGAGTCCCTGTATGATGAGATGCTTTAATAATAAAGATG 2355  
1716 GTCTGGAGATCTCTGTCTAGAGTCCCTGTATGATGAGATGCTTTAATAATAAAGATG 1775  
2356 TTTTGAT 2363  
1776 TTTTGAT 1783  
RESULT 9  
US-10-073-060-1  
; Sequence 1, Application US/10073060  
; Publication No. US20030049645A1  
; GENERAL INFORMATION:  
; APPLICANT: MU, David  
; APPLICANT: POWERS, Scott  
; TITLE OF INVENTION: Amplified Cancer Gene Hepsin  
; FILE REFERENCE: 38002-0024  
; CURRENT APPLICATION NUMBER: US/10/073,060  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-073-060-1  
Query Match 67.4%; Score 1593.6; DB 14; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
736 CCCCCAGGCTGCTGCTCCCTGCTCATCTCTCAAGATCCCAACCTGGCCAGAGGTCA 795  
157 CTTGAGAGCTCCGCCCCCACTGCTGAGACCCA-GGGTCCACCTTGCCCAAGAGGTCA 215  
796 GCCAGGAAATCAATTAACAAGAGCAGTACATGACGCAAGAGAGAGGTGCGGACTGTG 855  
216 GCCAGGAAATCAATTAACAAGAGCAGTACATGACGCAAGAGAGAGGTGCGGACTGTG 275  
856 CCATGCTCTTCAGACCCCAAGGTGAGCTTCTCACTGCGGGACCTTGCTATCTTGAACA 915  
276 CCATGCTCTTCAGACCCCAAGGTGAGCTTCTCACTGCGGGACCTTGCTATCTTGAACA 335  
916 GCCATCGGGGCGGATCTGTGGCCATTTGTGGCTTTCTCTCAGAGATGACAGAGACCG 975  
336 GCCATCGGGGCGGATCTGTGGCCATTTGTGGCTTTCTCTCAGAGATGACAGAGACCG 395  
976 CTGTACCCAGTGCAGGTCAAGCTTGCAGGACGCTCGGCTCAATGCTTTTGAACAAAGGAA 1035  
396 CTGTACCCAGTGCAGGTCAAGCTTGCAGGACGCTCGGCTCAATGCTTTTGAACAAAGGAA 455  
276 CCATGCTCTTCAGACCCCAAGGTGAGCTTCTCACTGCGGGACCTTGCTATCTTGAACA 335  
916 GCCATCGGGGCGGATCTGTGGCCATTTGTGGCTTTCTCTCAGAGATGACAGAGACCG 975  
336 GCCATCGGGGCGGATCTGTGGCCATTTGTGGCTTTCTCTCAGAGATGACAGAGACCG 395  
976 CTGTACCCAGTGCAGGTCAAGCTTGCAGGACGCTCGGCTCAATGCTTTTGAACAAAGGAA 1035  
396 CTGTACCCAGTGCAGGTCAAGCTTGCAGGACGCTCGGCTCAATGCTTTTGAACAAAGGAA 455  
456 GGGAGCTGGGGGCTGTGCTCTCTCGGCTCCAAACCCAGGAGGAGGAGCTCAGCTGC 515  
1036 GGGAGCTGGGGGCTGTGCTCTCTCGGCTCCAAACCCAGGAGGAGGAGCTCAGCTGC 1095  
1096 GAGAGATGGGCTTCTTCAGAGGCACTGACCTCCAGAGCTGAGCGTGGAGAGGCGGGG 1155  
516 GAGAGATGGGCTTCTTCAGAGGCACTGACCTCCAGAGCTGAGCGTGGAGAGGCGGGG 575  
1156 GCCAATGGACAGTGGGCTTCTTCTGTGTGAGACGAGGGGAGGCTGCCCAACCCAGAGG 1215  
576 GCCAATGGACAGTGGGCTTCTTCTGTGTGAGACGAGGGGAGGCTGCCCAACCCAGAGG 635  
1216 CTGCTGAGAGTCACTCTCGTGTGTGATTGGCCCAAGAGCGGCTTTCTTGGCGCATCTGC 1275  
636 CTGCTGAGAGTCACTCTCGTGTGTGATTGGCCCAAGAGCGGCTTTCTTGGCGCATCTGC 695  
1276 CAAGACTGTGGCCGAGGAGAGCTGCGGTGACCGCATCTGTGGAGGCGGGAGACACAGC 1335



|    |      |  |      |
|----|------|--|------|
| OY | 1336 | TCCCTGCTCTCCGGGGGACTGGGTGTGACAGCGCCCACTGCTTCCCGAGCCGGAAACCGG       | 1455 |
| Db | 816  | TCCCTGCTCTCCGGGGGACTGGGTGTGACAGCGCCCACTGCTTCCCGAGCCGGAAACCGG       | 875  |
| OY | 1456 | GTCCCTGTCGCGATGGCAGATGTTTGGCCGGTCCCGTGGCCAGGCTCTCCCCACGGTCTG       | 1515 |
| OY | 1516 | CAGCTGGGGGTGACAGGCTGTGTGTACACACGGGGGTATCTTCCCTTTCCGGGACCCCAAC      | 1575 |
| Db | 936  | CAGCTGGGGGTGACAGGCTGTGTGTACACACGGGGGTATCTTCCCTTTCCGGGACCCCAAC      | 995  |
| OY | 1576 | AGCGAGGAGAAACAGCAAGATATTGGCCCTGGGTCCACTCTCAGTCCCTGCCCCCTCA         | 1635 |
| Db | 996  | AGCGAGGAGAAACAGCAAGATATTGGCCCTGGGTCCACTCTCAGTCCCTGCCCCCTCA         | 1055 |
| OY | 1636 | GAATACATCCAGCCTGTGTGCTCCCGACGTGCCCGGACAGGCTGTGTGGATGGCAAGATC       | 1695 |
| Db | 1056 | GAATACATCCAGCCTGTGTGCTCCCGACGTGCCCGGACAGGCTGTGTGGATGGCAAGATC       | 1115 |
| OY | 1696 | TGTACCCGTGAACGGGTGTGGGGGCAACACGAGTACTATGGCCAAACAGGCCGGGGGTACTCCAG  | 1755 |
| Db | 1116 | TGTACCCGTGAACGGGTGTGGGGGCAACACGAGTACTATGGCCAAACAGGCCGGGGGTACTCCAG  | 1175 |
| OY | 1756 | GAGGCTCAGTCCCGCATTAATCAGCAATATGTCTGCAATGGCCGCTGACTTCTATGGAAAC      | 1815 |
| Db | 1176 | GAGGCTCAGTCCCGCATTAATCAGCAATATGTCTGCAATGGCCGCTGACTTCTATGGAAAC      | 1235 |
| OY | 1816 | CAGATCAAGCCCAAGATATGTTGTGTGTGCTGGTCAACCCGAGGGGTGCAATTGATGCTGCAAG   | 1875 |
| Db | 1236 | CAGATCAAGCCCAAGATATGTTGTGTGTGCTGGTCAACCCGAGGGGTGCAATTGATGCTGCAAG   | 1295 |
| OY | 1876 | GGCCACAGCCGGTGGTCCCTTTGTGTGTGTGAGGACACAGACTCTTCGGAGCGCAAGTTGGCCGG  | 1935 |
| Db | 1296 | GGCCACAGCCGGTGGTCCCTTTGTGTGTGTGAGGACACAGACTCTTCGGAGCGCAAGTTGGCCGG  | 1355 |
| OY | 1936 | CTGTGTGGCATTTGTGTGATGTGGGGGCACTGGCTGTGGCCCTGGGCCAGAAAGCCAGGGGTCTAC | 1995 |
| Db | 1356 | CTGTGTGGCATTTGTGTGATGTGGGGGCACTGGCTGTGGCCCTGGGCCAGAAAGCCAGGGGTCTAC | 1415 |
| OY | 1996 | ACCAAAAGTCAGTGAATTCGCGGAGTGGATCTTCCAGGCCATTAAGACTCACTCCGAAGCC      | 2055 |
| Db | 1416 | ACCAAAAGTCAGTGAATTCGCGGAGTGGATCTTCCAGGCCATTAAGACTCACTCCGAAGCC      | 1475 |
| OY | 2056 | AGCGGCATGTGTGAACCCAGCTCTGACCCGGTGGCTTCTGGCTGCGCAGGCTCCAGGGCCCGA    | 2115 |
| Db | 1476 | AGCGGCATGTGTGAACCCAGCTCTGACCCGGTGGCTTCTGGCTGCGCAGGCTCCAGGGCCCGA    | 1535 |
| OY | 2116 | GGTGATCCCGGGTGGTGGGATCCAGCTGGGGCGAGAGATGGGACGTTTTTCTTCTTGAGGAC     | 2175 |
| Db | 1536 | GGTGATCCCGGGTGGTGGGATCCAGCTGGGGCGAGAGATGGGACGTTTTTCTTCTTGAGGAC     | 1595 |
| OY | 2176 | CGGTCCACAGGTCCAAAGGACACCTCCTCCAGGGGTCTCTTCCACAGTGGCGGGCC           | 2235 |
| Db | 1596 | CGGTCCACAGGTCCAAAGGACACCTCCTCCAGGGGTCTCTTCCACAGTGGCGGGCC           | 1655 |
| OY | 2236 | ACTCAGCCCCGAGACACACCCCACTCAACCTCTCTGACCCCCCATGTAAATATTGTTCTGCT     | 2295 |
| Db | 1656 | ACTCAGCCCCGAGACACACCCCACTCAACCTCTCTGACCCCCCATGTAAATATTGTTCTGCT     | 1715 |
| OY | 2296 | GTCTGGGAGCTCTGTCTAGTGGCCCTGTATGATGGAGTGTCTTTAAATAATAAAGATGG        | 2355 |
| Db | 1716 | GTCTGGGAGCTCTGTCTAGTGGCCCTGTATGATGGAGTGTCTTTAAATAATAAAGATGG        | 1775 |
| OY | 2356 | TTTTTGATT 2363   |      |
| Db | 1776 | TTTTTGATT 1783   |      |





Db 696 CAAGACTGTGCGCGAGAGAGCTGCCCGTGGACCCCATCTGTGGAGGCCGGAGACACACAGC 755  
Qy 1336 TTGGGCGCGGTGGCCCTGTGCAATCAGCTTGCCTATGATGAGACACACCTCTGTGGGAG 1395  
Db 756 TTGGGCGCGGTGGCCCTGTGCAATCAGCTTGCCTATGATGAGACACACCTCTGTGGGAG 815  
Qy 1396 TCCCGCTCTCCGGGGAGCTGGGTGTGTCAGCCGCGCCACTGCTTCCCGAGGCGAAACGG 1455  
Db 816 TCCCTGCTCTCCGGGGAGCTGGGTGTGTCAGCCGCGCCACTGCTTCCCGAGGCGAAACGG 875  
Qy 1456 GTCTGTCTCCGATGCGAGATGTTTCCCGGTGGCCGTGGCCAGAGCTCTCCCAAGGCTGTG 1515  
Db 876 GTCTGTCTCCGATGCGAGATGTTTCCCGGTGGCCGTGGCCAGAGCTCTCTCCCAAGGCTGTG 935  
Qy 1516 CAGCTGGGGGTGTCAGGCTGTGTGTCACACGGGGGCTATCTTCCCTTCCGGAGCCCAAC 1575  
Db 936 CAGCTGGGGGTGTCAGGCTGTGTGTCACACGGGGGCTATCTTCCCTTCCGGAGCCCAAC 995  
Qy 1576 AGCGAGAGAAACAGACGATTTTCCCTGTGTCACCTCTCCAGTCCCTGTGCTCCAC 1635  
Db 996 AGCGAGAGAAACAGACGATTTTCCCTGTGTCACCTCTCCAGTCCCTGTGCTCCAC 1055  
Qy 1636 GAATATATCCAGCTGTGTGCTCCAGCTGCGGCGAGCCCTGTGATGGCAAGATC 1695  
Db 1056 GAATATATCCAGCTGTGTGCTCCAGCTGCGGCGAGCCCTGTGATGGCAAGATC 1115  
Qy 1696 TGTATCCGTGACGGGCTGGGGCAACCGAGTACTATGGCCAAACAGCCGGGGTACTCCAG 1755  
Db 1116 TGTATCCGTGACGGGCTGGGGCAACCGAGTACTATGGCCAAACAGCCGGGGTACTCCAG 1175  
Qy 1756 GAGGTCCAGTCCCATATATGACAAATGATGTCTGCAATGGGGCTGATCTTCTATGGAAC 1815  
Db 1176 GAGGTCCAGTCCCATATATGACAAATGATGTCTGCAATGGGGCTGATCTTCTATGGAAC 1235  
Qy 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGCTGCGCAG 1875  
Db 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGCTGCGCAG 1295  
Qy 1876 GGCGCAGCGGTGTGCTCTTGTGTGTGAGAGACAGCATCTCTCGAGCCCAAGTTGGGG 1935  
Db 1296 GGCGCAGCGGTGTGCTCTTGTGTGTGAGAGACAGCATCTCTCGAGCCCAAGTTGGGG 1355  
Qy 1936 CTGTGTGCTGATGTGAGTTGGGGCACTGGCTGTGCTGCGCCAGAACCCAGGCTGTAC 1995  
Db 1356 CTGTGTGCTGATGTGAGTTGGGGCACTGGCTGTGCTGCGCCAGAACCCAGGCTGTAC 1415  
Qy 1996 ACCAAGTCAGTGAATTCGGGAGTGAATCTTCCAGGCAATTAAGATCACTCCGAGAGCC 2055  
Db 1416 ACCAAGTCAGTGAATTCGGGAGTGAATCTTCCAGGCAATTAAGATCACTCCGAGAGCC 1475  
Qy 2056 AGCGGATGTGACCCAGCTTGAACCGGTGTGCTTCTCGTGGCAGCCCTCCAGAGGCCGA 2115  
Db 1476 AGCGGATGTGACCCAGCTTGAACCGGTGTGCTTCTCGTGGCAGCCCTCCAGAGGCCGA 1535  
Qy 2116 GGTGATCCCGGTGTGAGATCAAGCTGTGGCCGAGAGATGGAGACGTTTCTTCTTGGGCC 2175  
Db 1536 GGTGATCCCGGTGTGAGATCAAGCTGTGGCCGAGAGATGGAGACGTTTCTTCTTGGGCC 1595  
Qy 2176 CGGTTCACAGGTCGAAGACACCCCTCCAGGGGTCTCTTCCAGAGTGGGGGGGCC 2235  
Db 1596 CGGTTCACAGGTCGAAGACACCCCTCCAGGGGTCTCTTCCAGAGTGGGGGGGCC 1655  
Qy 2236 ACTCAGCCCGAGACCAACCACTCACTCTGACCCCAATGTAATATTTGTTCTGCT 2295  
Db 1656 ACTCAGCCCGAGACCAACCACTCACTCTGACCCCAATGTAATATTTGTTCTGCT 1715  
Qy 2296 GTCTGGGACTCTGTCTAGGTGCTGCTGATGATGGAGTCTTTAAATAAATGAATGG 2355  
Db 1716 GTCTGGGACTCTGTCTAGGTGCTGCTGATGATGGAGTCTTTAAATAAATGAATGG 1775  
Qy 2356 TTTTGATT 2363  
|||||

Db 1776 TTTTGATT 1783  
RESULT 13  
US-10-295-027-1160  
; Sequence 1160, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afari, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezl, Peter A.  
; APPLICANT: Mack, David H.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; FILE REFERENCE: 018501-012500US  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: US 60/340,376  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: US 60/347,211  
; PRIOR FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US 60/347,349  
; PRIOR FILING DATE: 2002-01-10  
; PRIOR APPLICATION NUMBER: US 60/355,250  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US 60/356,714  
; PRIOR FILING DATE: 2002-02-13  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1386  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1160  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-295-027-1160  
Query Match 67.4%; Score 1593.6; DB 17; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
Qy 736 CCCCCAGCCCTGCTCCCGCTCCATCTCTCAGAGGTCCACCTTGGCCAGAGGTCA 795  
Db 157 CCTGAGGCTCCCGCCCACTGCTGAGACCCA-GGGTCCACCTTGGCCAGAGGTCA 215  
Qy 796 GCCAGGATCATTAACAAGAGAGTGAATGGCCGAGAGAGGTGGCCGAGCTGTG 855  
Db 216 GCCAGGATCATTAACAAGAGAGTGAATGGCCGAGAGAGGTGGCCGAGCTGTG 275  
Qy 856 CCATGCTGCTCAGAGCCCAAGTGGAGCTCTCAGTGGGGACCTTGTACTTCTGACA 915  
Db 276 CCATGCTGCTCAGAGCCCAAGTGGAGCTCTCAGTGGGGACCTTGTACTTCTGACA 335  
Qy 916 GCCATCGGGGGGAGTCTTGGGCTTGTGTGCTTCTCTCAGAGTGAACAGAGCCG 975  
Db 336 GCCATCGGGGGGAGTCTTGGGCTTGTGTGCTTCTCTCAGAGTGAACAGAGCCG 395  
Qy 976 CTGTATCCAGTGCAGTCACTGTGCGAGCGCTCGGCTCATGTCTTTGACAAACGGA 1035  
|||||

Db 396 CTGATCCAGGTGACAGGTCTGCTGGGACGCTGGCTCATGTCTTTTGACAAAGCCGAA 455  
Oy 1036 GGAAGCTGGCGGCTGTGTCTCTCGCGCTCAACGCCAGGGTAAGCCGACTCAGCTGC 1095  
Db 456 GGGAGCTGGCGGCTGTGTCTCTCGCGCTCAACGCCAGGGTAAGCCGACTCAGCTGC 515  
Oy 1096 GAGGAGATGGGCTTCTCAAGGCACTGACCCACTCCGAGCTGGAGCGTGGAAACGGCGGGC 1155  
Db 516 GAGGAGATGGGCTTCTCAAGGCACTGACCCACTCCGAGCTGGAGCGTGGAAACGGCGGGC 575  
Oy 1156 GCCAATGGCAGCTGGGCTTCTGTGTGAGCAAGAGGAGAGCTGCCCCACACCCAGAG 1215  
Db 576 GCCAATGGCAGCTGGGCTTCTGTGTGAGCAAGAGGAGAGCTGCCCCACACCCAGAG 635  
Oy 1216 CTGCTGAGAGTCTATCTCGGTGTGATTTGCCCAGAGGCGTTTCTTGGCCCATCTGC 1275  
Db 636 CTGCTGAGAGTCTATCTCGGTGTGATTTGCCCAGAGGCGTTTCTTGGCCCATCTGC 695  
Oy 1276 CAAGACTGTGGCCGAGGAGCTGCGGTGACGCGATCGTGGAGCGCGGGACACACG 1335  
Db 696 CAAGACTGTGGCCGAGGAGCTGCGGTGACGCGATCGTGGAGCGCGGGACACACG 755  
Oy 1336 TTGGGCGCGGTGCGGTGCAAGTCAAGCTTGTATGATGAGACACCTCTGTGGGGA 1395  
Db 756 TTGGGCGCGGTGCGGTGCAAGTCAAGCTTGTATGATGAGACACCTCTGTGGGGA 815  
Oy 1396 TTCCCTGCTCTCGGGGAGCTGGGTGTGACAGCCGCCACTGCTTCCGGAGCGAAACGG 1455  
Db 816 TTCCCTGCTCTCGGGGAGCTGGGTGTGACAGCCGCCACTGCTTCCGGAGCGAAACGG 875  
Oy 1456 GTCCGTGTCCGATGGCGAGTGTGTTGGCGGTGCGGTGCGGCGCAAGGCTCTCCCAAGGTG 1515  
Db 876 GTCCGTGTCCGATGGCGAGTGTGTTGGCGGTGCGGTGCGGCGCAAGGCTCTCCCAAGGTG 935  
Oy 1516 CAGCTGGGGGTGACAGGCTGTGTATCCACAGGGGCTATCTTCCCTTTCGGGACCCCAAC 1575  
Db 936 CAGCTGGGGGTGACAGGCTGTGTATCCACAGGGGCTATCTTCCCTTTCGGGACCCCAAC 995  
Oy 1576 AGCGAGGAGAAACAGCAAGATATGCTGCTGCTCAAGTCCCTGCTCCCTCTCA 1635  
Db 996 AGCGAGGAGAAACAGCAAGATATGCTGCTGCTCAAGTCCCTGCTCCCTCTCA 1055  
Oy 1636 GAATACATCCAGCTGTGTGCTCCAGCTGCGGGCCAGGCGCTGTGATGGCAAGATC 1695  
Db 1056 GAATACATCCAGCTGTGTGCTCCAGCTGCGGGCCAGGCGCTGTGATGGCAAGATC 1115  
Oy 1696 TGTACCTGTGACGGGCTGGGGCAACGCGAGTACTATGACCAAGGCGGGGTACTCCAG 1755  
Db 1116 TGTACCTGTGACGGGCTGGGGCAACGCGAGTACTATGACCAAGGCGGGGTACTCCAG 1175  
Oy 1756 GAGGCTCGAGTCCCATATATGACCAATATGTCTGCAATGGCGCTGACTTCTTATGAAAC 1815  
Db 1176 GAGGCTCGAGTCCCATATATGACCAATATGTCTGCAATGGCGCTGACTTCTTATGAAAC 1235  
Oy 1816 CAGATCAAGCCCAAGATGTCTGTGCTGTGAGGACAGCATCTCTGGAACGCACTTGGCGG 1875  
Db 1236 CAGATCAAGCCCAAGATGTCTGTGCTGTGAGGACAGCATCTCTGGAACGCACTTGGCGG 1295  
Oy 1876 GGGCAGCAGCGGTGCTCTTGTGTGTGAGGACAGCATCTCTGGAACGCACTTGGCGG 1935  
Db 1296 GGGCAGCAGCGGTGCTCTTGTGTGTGAGGACAGCATCTCTGGAACGCACTTGGCGG 1355  
Oy 1936 CTGTGTGGCATTTGTGAGTGGGGGCACTGGCTGTGCTTGGCGGCAAGAGCGGTCTAC 1995  
Db 1356 CTGTGTGGCATTTGTGAGTGGGGGCACTGGCTGTGCTTGGCGGCAAGAGCGGTCTAC 1415  
Oy 1996 ACCAAAGTCAGTGAATTCGGGAGTGTGATCTTCCAGGCGATTAAGTCACTCCGAAGCC 2055  
Db 1416 ACCAAAGTCAGTGAATTCGGGAGTGTGATCTTCCAGGCGATTAAGTCACTCCGAAGCC 1475  
Oy 2056 AGCGGATGTGACCCAGCTCTGACCGGTGCTTCTCGCTGCGGACGCTCCAGGAGCCGA 2115

Db 1476 AGCGGATGTGACCCAGCTCTGACCGGTGCTTCTCGTGGCAGGCTCCAGGAGCCGA 1535  
Oy 2116 GGTGATTCGGGTGGTGGGATTCACAGCTGGGCGGAGATGGGAGCTTTTCTTGTGGGC 2175  
Db 1536 GGTGATTCGGGTGGTGGGATTCACAGCTGGGCGGAGATGGGAGCTTTTCTTGTGGGC 1595  
Oy 2176 CGGTCCAGAGTCCAAAGACACCTCTCCAGGGTCTCTCTTCCACAGTGGCGGGCC 2235  
Db 1596 CGGTCCAGAGTCCAAAGACACCTCTCCAGGGTCTCTCTTCCACAGTGGCGGGCC 1655  
Oy 2236 ACTCAGCTCCGAGACCAACCACTCACTCTGACCCCACTGTAATATTTGCT 2295  
Db 1656 ACTCAGCTCCGAGACCAACCACTCACTCTGACCCCACTGTAATATTTGCT 1715  
Oy 2296 GTCTGGAGCTCTGTCTAGTGGCGGCTGATGATGGGATGCTTTAATAATAAGATG 2355  
Db 1716 GTCTGGAGCTCTGTCTAGTGGCGGCTGATGATGGGATGCTTTAATAATAAGATG 1775  
Oy 2356 TTTTGATT 2363  
Db 1776 TTTTGATT 1783

RESULT 14  
US-10-156-214A-34  
; Sequence 34, Application US/10156214A  
; Publication No. US20040001801A1  
; GENERAL INFORMATION:  
; APPLICANT: Edwin L. Madison  
; APPLICANT: Joseph Edward Semple  
; APPLICANT: George P. Vlasuk  
; APPLICANT: Scott Jeffrey Kemp  
; APPLICANT: Mallareddy Komandla  
; APPLICANT: Daniel Vanma Stiev  
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic U  
; FILE REFERENCE: 24745-1611  
; CURRENT APPLICATION NUMBER: US/10/156,214A  
; NUMBER OF SEQ ID NOS: 611  
; SOFTWARE: FastSeq For Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; NAME/KEY: CDS  
; LOCATION: (246)...(1499)  
; OTHER INFORMATION: Nucleic acid encoding human hepsin  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: GenBank M18930  
; DATABASE ENTRY DATE: 1993-06-11  
US-10-156-214A-34

Query Match 67.4%; Score 1593.6; DB 17; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Db 736 CCCCCAGGCTTGGCTTCCGCTTCATCTCTCAAGTCCCACTTGGAGGAGGAGTCA 795  
Oy 157 CTTGAGAGCTTCCGCCCACTGCTGAGACCCA-GGTTCCACCTTGGCCAGAGGATCA 215  
Db 796 GCCAGGAAATCAATTAACAAGGCAATGAGCATGGCCGAGAAGAGGGTGGCGGACTGTG 855  
Oy 216 GCCAGGAAATCAATTAACAAGGCAATGAGCATGGCCGAGAAGAGGGTGGCGGACTGTG 275  
Db 856 CCATGCTCTCAGACCCCAAGTGGCAGCTCTCACTGCGGGGACCCCTGCTACTTCTGACA 915  
Oy 276 CCATGCTCTCAGACCCCAAGTGGCAGCTCTCACTGCGGGGACCCCTGCTACTTCTGACA 335  
Db 916 GCCATGCGGGCGGCGATCTTGGGCGATTTGGCTGTTCTCTCAGAGATGACAGAGACCG 975  
Oy 336 GCCATGCGGGCGGCGATCTTGGGCGATTTGGCTGTTCTCTCAGAGATGACAGAGACCG 395

```

QY 976 CTGTACCAAGTCAGAGTCCTGCGGAGCGCTCGGCTCATGCTCTTTGACAAGACGAA 1035
DB 396 CTGTACCAAGTCAGAGTCCTGCGGAGCGCTCGGCTCATGCTCTTTGACAAGACGAA 455
QY 1036 GGGACGTGCGGCTGCTGCTCTGCGGCTCAACCGCAGGGTACGCGGACCTGAGCTGC 1095
DB 456 GGGACGTGCGGCTGCTGCTCTGCGGCTCAACCGCAGGGTACGCGGACCTGAGCTGC 515
QY 1096 GAGGAGATGAGCTTCTTCAAGGACCTGACCCATCTCCAGCTGAGAGTGGAAAGCGGAGC 1155
DB 516 GAGGAGATGAGCTTCTTCAAGGACCTGACCCATCTCCAGCTGAGAGTGGAAAGCGGAGC 575
QY 1156 GCCATGAGACGTCGGGCTTCTTCTGTGTGAGAGAGGGAGGCTGCGGACCAACCGAGAG 1215
DB 576 GCCATGAGACGTCGGGCTTCTTCTGTGTGAGAGAGGGAGGCTGCGGACCAACCGAGAG 635
QY 1216 CTGCTGAGAGTATCTCCGTGTGTGATTGCGCCAGAGGCGCTTCTTGGCGCGCATCTGC 1275
DB 636 CTGCTGAGAGTATCTCCGTGTGTGATTGCGCCAGAGGCGCTTCTTGGCGCGCATCTGC 695
QY 1276 CAAGACTGTGCGCGAGAGAGCTGCGCGGTGAGCGGATCTGTGAGAGCGCGGAGACACAGC 1335
DB 696 CAAGACTGTGCGCGAGAGAGCTGCGCGGTGAGCGGATCTGTGAGAGCGCGGAGACACAGC 755
QY 1336 TTGGGCGGCTGCGCGGTGAGAGCTGCGCGGTGAGCGGATCTGTGAGAGCGCGGAGACACAGC 1395
DB 756 TTGGGCGGCTGCGCGGTGAGAGCTGCGCGGTGAGCGGATCTGTGAGAGCGCGGAGACACAGC 815
QY 1396 TCCCTGCTCTCCGCGGAGCTGAGTGTGACAGCGCGGCTTCTTCCGAGAGCGGAGACCGG 1455
DB 816 TCCCTGCTCTCCGCGGAGCTGAGTGTGACAGCGCGGCTTCTTCCGAGAGCGGAGACCGG 875
QY 1456 GTCTGTCCGATGCGAGAGTGTGCGCGGTGCGCGGTGCGCGGCTTCTTCCGAGAGCTGTG 1515
DB 876 GTCTGTCCGATGCGAGAGTGTGCGCGGTGCGCGGTGCGCGGCTTCTTCCGAGAGCTGTG 935
QY 1516 CAGCTGGGGGTGCGAGGCTGTGTCTACACAGGGGGCTTCTTCCGAGAGCGGAGACCGG 1575
DB 936 CAGCTGGGGGTGCGAGGCTGTGTCTACACAGGGGGCTTCTTCCGAGAGCGGAGACCGG 995
QY 1576 AGCGAGAGAGACAGAGAGATTTGCGCTGTCACCTCTCCAGTCCCGCTGACCTGAC 1635
DB 996 AGCGAGAGAGAGACAGAGATTTGCGCTGTCACCTCTCCAGTCCCGCTGACCTGAC 1055
QY 1636 GAATACATCCAGCGCTGTGTGCTCCAGCTGCGGCGGAGCGGCTGTGTGAGTGGAGAGATC 1695
DB 1056 GAATACATCCAGCGCTGTGTGCTCCAGCTGCGGCGGAGCGGCTGTGTGAGTGGAGAGATC 1115
QY 1696 TGTACCGTGAAGGCTGTGGGCAACAGCATCTACTATGCGCAACAGGCGGGGTACTCCAG 1755
DB 1116 TGTACCGTGAAGGCTGTGGGCAACAGCATCTACTATGCGCAACAGGCGGGGTACTCCAG 1175
QY 1756 GAGGCTGAGTCCCATATATGCAATGATGTCTGCAATGAGGCTGACTTCTATGGAAC 1815
DB 1176 GAGGCTGAGTCCCATATATGCAATGATGTCTGCAATGAGGCTGACTTCTATGGAAC 1235
QY 1816 CAGATCAAGCCCAAGATGTTCTGTCTGTGCTGCTACCCGAGAGGTGAGATGATGCTGAG 1875
DB 1236 CAGATCAAGCCCAAGATGTTCTGTCTGTGCTGCTACCCGAGAGGTGAGATGATGCTGAG 1295
QY 1876 GGGCAGACGGGTGCTCTTGTGTGTGAGAGACAGCATCTCTGAGAGCGGCTGTGGGG 1935
DB 1296 GGGCAGACGGGTGCTCTTGTGTGTGAGAGACAGCATCTCTGAGAGCGGCTGTGGGG 1355
QY 1936 CTGTGTGAGTGTGAGTGTGGGCACTGGCTGTGCTGCGCCAGAGCGGAGCGCTTAC 1995
DB 1356 CTGTGTGAGTGTGAGTGTGGGCACTGGCTGTGCTGCGCCAGAGCGGAGCGCTTAC 1415
QY 1996 ACCAAGTCACTGACTTCCGAGAGTGTGATCTTCCAGGCGATTAAGACTCACTCCGAGCC 2055
DB 1416 ACCAAGTCACTGACTTCCGAGAGTGTGATCTTCCAGGCGATTAAGACTCACTCCGAGCC 1475

```

```

QY 2056 AGCGCATGTGATACCCAGCTCTGACCGGTGAGCTTCTGCTGCGAGCTCCAGGCGCGGA 2115
DB 1476 AGCGCATGTGATACCCAGCTCTGACCGGTGAGCTTCTGCTGCGAGCTCCAGGCGCGGA 1535
QY 2116 GGTATCCCGGTGTGAGATCCAGCTGAGCGGCGGAGATGGAGCGTTTCTTCTTGGGCGC 2175
DB 1536 GGTATCCCGGTGTGAGATCCAGCTGAGCGGCGGAGATGGAGCGTTTCTTCTTGGGCGC 1595
QY 2176 CGGTTCACAGTCCCAAGAGACACCTTCTCCAGAGGTCTCTTCTTCCACAGTGGCGGCGC 2235
DB 1596 CGGTTCACAGTCCCAAGAGACACCTTCTCCAGAGGTCTCTTCTTCCACAGTGGCGGCGC 1655
QY 2236 ACTGAGCGCGGAGACCAACCACTGACCTCTCTGAGACCCCATGTAATATTTGTTCTGCT 2295
DB 1656 ACTGAGCGCGGAGACCAACCACTGACCTCTCTGAGACCCCATGTAATATTTGTTCTGCT 1715
QY 2296 GTCTGGAGCTCTGTCTAGTGTGCGGCTGATGATGGATGCTTTTAAATTAATTAAGATGG 2355
DB 1716 GTCTGGAGCTCTGTCTAGTGTGCGGCTGATGATGGATGCTTTTAAATTAATTAAGATGG 1775
QY 2356 TTTTGATT 2363
DB 1776 TTTTGATT 1783

RESULT 15
US-10-652-993-188
; Sequence 188, Application US/10652993
; Publication No. US20040166117A1
; GENERAL INFORMATION:
; APPLICANT: Cannon, Martin J.
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Santin, Alessandro
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
; FILE REFERENCE: D6223CIP/A/D/CIP4
; CURRENT FILING DATE: US/10/652, 993
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: 10/135, 795
; PRIOR FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 195
; SEQ ID NO 188
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length cDNA of hepsin
US-10-652-993-188

Query Match 67.4%; Score 1593.6; DB 19; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCAGGCGCTGCTCCCTCCCTCATCTCTACAGGTGCCACCTGCGGAGAGTCA 795
DB 157 CTTGAGGCTCGGCCCACTGCTGAGACCCCA-GGGTCCACCTGCGGAGAGTCA 215
QY 796 GCCAGGATCATTAACAAGGAGTGAATGCGCAGAGAGAGGAGGTCGAGCTGTG 855
DB 216 GCCAGGATCATTAACAAGGAGTGAATGCGCAGAGAGAGGAGGTCGAGCTGTG 275
QY 856 CCATGCTGCTCAAGACCAAGTGGAGCTCACTGCGGAGACCTGCTACTTGTGCA 915
DB 276 CCATGCTGCTCAAGACCAAGTGGAGCTCACTGCGGAGACCTGCTACTTGTGCA 335
QY 916 GCCATGCGGAGCGGATCTGCGGATGCTGTGCTGCTCTCCTCAGAGTGAACAGAGCGG 975
DB 336 GCCATGCGGAGCGGATCTGCGGATGCTGTGCTGCTCTCCTCAGAGTGAACAGAGCGG 395
QY 976 CTGTACCAAGTCACTGAGTCACTTCTGCGAGCGCTGCTCATGTGCTTTGACAAGCGAA 1035
DB 396 CTGTACCAAGTCACTGAGTCACTTCTGCGAGCGCTGCTCATGTGCTTTGACAAGCGAA 455
QY 1036 GGGACGTGCGGCTGCTGCTCTCGGCTCAACCGCAGGGTACCGGAGCTACGCTGC 1095

```

Db 456 GGGACGTGGCCGGCTGTGTGCTCTCCGGCTCCAAACGCAAGGATACCGGACTCAGCTGC 515  
Qy 1096 GAGAGATGGGCTTCTCTCAAGGCACTGACCACTCCGACTGGAAGTGCACAAGCGGCGGC 1155  
Db 516 GAGAGATGGGCTTCTCTCAAGGCACTGACCACTCCGACTGGAAGTGCACAAGCGGCGGC 575  
Qy 1156 GCCAATGGCAGGTGGGGCTTCTGTGTGGAACGAGGGAGAGCTGCCCCACACCCAGAG 1215  
Db 576 GCCAATGGCAGGTGGGGCTTCTGTGTGGAACGAGGGAGAGCTGCCCCACACCCAGAG 635  
Qy 1216 CTGCTGAGGATCATCTCCGTGTGTGATTTGCCAGAGGCGTTTCTTGGCCGCACTGCG 1275  
Db 636 CTGCTGAGGATCATCTCCGTGTGTGATTTGCCAGAGGCGTTTCTTGGCCGCACTGCG 695  
Qy 1276 CAAGACTGTGGCCGCAAGAACCTGCCGTGACCGCATGTGGAGGCGCGGACAACAGC 1335  
Db 696 CAAGACTGTGGCCGCAAGAACCTGCCGTGACCGCATGTGGAGGCGCGGACAACAGC 755  
Qy 1336 TTGGGCGGTTGGCCGTTGGCAAGTCAAGCTTGATGATGAGGACACCTGTGTGGGGGA 1395  
Db 756 TTGGGCGGTTGGCCGTTGGCAAGTCAAGCTTGATGATGAGGACACCTGTGTGGGGGA 815  
Qy 1396 TCCCTGTCTCTCCGGGGACTGGGTGTGACAGCCGCGCACTGCTTCCCGAGCGGAACCGG 1455  
Db 816 TCCCTGTCTCTCCGGGGACTGGGTGTGACAGCCGCGCACTGCTTCCCGAGCGGAACCGG 875  
Qy 1456 GTCTGTCTCCGATGGAGGATGTTTGCCTGTGCTGAGCCCAAGGCTCTCCCAAGGCTG 1515  
Db 876 GTCTGTCTCCGATGGAGGATGTTTGCCTGTGCTGAGCCCAAGGCTCTCCCAAGGCTG 935  
Qy 1516 CAGCTGGGGGGTGCAGGCTGTGTGTATCCAGCGGGGCTATCTTCCCTTTCGGGACCCCAAC 1575  
Db 936 CAGCTGGGGGGTGCAGGCTGTGTGTATCCAGCGGGGCTATCTTCCCTTTCGGGACCCCAAC 995  
Qy 1576 AGCGAGGAGAACAGCAACGATATTTGCCGTGTCCACTCTCCAGTCCCTGCCCCCTCACA 1635  
Db 996 AGCGAGGAGAACAGCAACGATATTTGCCGTGTCCACTCTCCAGTCCCTGCCCCCTCACA 1055  
Qy 1636 GAATATCATCCAGCTGTGTGCTCCCAAGCTGCCGAGCGCTGTGTGATGGAAGATC 1695  
Db 1056 GAATATCATCCAGCTGTGTGCTCCCAAGCTGCCGAGCGCTGTGTGATGGAAGATC 1115  
Qy 1696 TGTACCGTGAAGGGCTGGGGGCAACGAGTACTATGGCCAAAGGCCGGGGTACTCCAG 1755  
Db 1116 TGTACCGTGAAGGGCTGGGGGCAACGAGTACTATGGCCAAAGGCCGGGGTACTCCAG 1175  
Qy 1756 GAGGCTCGAGTCCCATATATCAGCAATGATGTGCAATGGCGCTGACTTCTATGGAAC 1815  
Db 1176 GAGGCTCGAGTCCCATATATCAGCAATGATGTGCAATGGCGCTGACTTCTATGGAAC 1235  
Qy 1816 CAGATCAAGGCCCAAGATGTTCTGTGTGTGCTAACCCCCGAGGGTGGATTGATGCCCTCAG 1875  
Db 1236 CAGATCAAGGCCCAAGATGTTCTGTGTGTGCTAACCCCCGAGGGTGGATTGATGCCCTCAG 1295  
Qy 1876 GGCGCAGCGGTGGTCCCTTGTGTGTGAGGACAGCATCTCCGAGCGCAGGTTGGCGG 1935  
Db 1296 GGCGCAGCGGTGGTCCCTTGTGTGTGAGGACAGCATCTCCGAGCGCAGGTTGGCGG 1355  
Qy 1936 CTGTGTGACATTTGATGATTTGGGCACTGGCTGTGCCCTGGCCAGAAACAGGCGCTTAC 1995  
Db 1356 CTGTGTGACATTTGATGATTTGGGCACTGGCTGTGCCCTGGCCAGAAACAGGCGCTTAC 1415  
Qy 1996 ACCAAAGTCAAGTCTTCCGGAGTGGATCTTCCAGGCCATTAAGACTCACTCCGAAGCC 2055  
Db 1416 ACCAAAGTCAAGTCTTCCGGAGTGGATCTTCCAGGCCATTAAGACTCACTCCGAAGCC 1475  
Qy 2056 AGCGGCAATGTGACCCAGCTGACCCGGTGGCTTCTGCTGCGGAGCTCCAGGGCCCGA 2115  
Db 1476 AGCGGCAATGTGACCCAGCTGACCCGGTGGCTTCTGCTGCGGAGCTCCAGGGCCCGA 1535  
Qy 2116 GGTGATCCCGGTGTGGATCCAGCTGGCCGAGATGGAGCGTTTCTTCTTGGGCGC 2175

Db 1536 GGTGATCCCGGTGTGGATCCAGCTGGGCCGAGATGGGACGTTTTCTTCTTGGGCGC 1595  
Qy 2176 GGTGATCCCGGTGTGGATCCAGCTGGGCCGAGATGGGACGTTTTCTTCTTGGGCGC 2235  
Db 1596 GGTGATCCCGGTGTGGATCCAGCTGGGCCGAGATGGGACGTTTTCTTCTTGGGCGC 1655  
Qy 2236 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACCCCGATGTAATATTTGTTGCT 2295  
Db 1656 ACTGAGCCCGGAGACCAACCACTCAACCTCTGACCCCGATGTAATATTTGTTGCT 1715  
Qy 2296 GTCTGGGACTCTGTGTGATGCTGCTGATGATGATGATGATGATGATGATGATGATG 2355  
Db 1716 GTCTGGGACTCTGTGTGATGCTGCTGATGATGATGATGATGATGATGATGATGATG 1775  
Qy 2356 TTTTGATT 2363  
Db 1776 TTTTGATT 1783

Search completed: August 17, 2005, 17:58:19  
Job time : 4019 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 01:25:31 / Search time 16703 Seconds

(without alignment)  
6855.037 Million cell updates/sec

Title: US-10-054-498-1  
Perfect score: 2363  
Sequence: 1 tcgagccgccttcacagga.....taataaagatgtttgattc 2363

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenBank1:  
1: gb ha:  
2: gb hrg:  
3: gb hn:  
4: gb om:  
5: gb ov:  
6: gb pat:  
7: gb ph:  
8: gb pl:  
9: gb pr:  
10: gb ro:  
11: gb stb:  
12: gb sy:  
13: gb un:  
14: gb vt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | ID       | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 2363   | 100.0       | 2363   | AR220138 | AR220138 Sequence  |
| 2          | 2363   | 100.0       | 2363   | AX336397 | AX336397 Sequence  |
| 3          | 2363   | 100.0       | 2363   | AX411046 | AX411046 Sequence  |
| 4          | 2363   | 100.0       | 2363   | AX511632 | AX511632 Sequence  |
| 5          | 2363   | 100.0       | 2363   | HSHEPSH  | HSHEPSH Sequence   |
| 6          | 1593.6 | 67.4        | 1783   | CO718142 | CO718142 Sequence  |
| 7          | 1593.6 | 67.4        | 1783   | AR280046 | AR280046 Sequence  |
| 8          | 1593.6 | 67.4        | 1783   | AX207963 | AX207963 Sequence  |
| 9          | 1593.6 | 67.4        | 1783   | HUMHPSNA | HUMHPSNA Sequence  |
| 10         | 1592   | 67.4        | 1761   | BC025716 | BC025716 Homo sapi |
| 11         | 1405   | 59.5        | 1769   | CO771486 | CO771486 Sequence  |
| 12         | 1269   | 53.7        | 1615   | AR255885 | AR255885 Sequence  |
| 13         | 1199   | 50.7        | 1199   | HSHEPSH  | HSHEPSH Sequence   |
| 14         | 1023.6 | 43.3        | 1739   | RNHEPSL  | RNHEPSL Sequence   |
| 15         | 1022.2 | 43.3        | 1781   | AF030065 | AF030065 Mus muscu |
| 16         | 1008.2 | 42.7        | 1605   | AR085656 | AR085656 Sequence  |
| 17         | 906.2  | 38.3        | 1887   | AY234104 | AY234104 Mus muscu |
| 18         | 858.4  | 36.3        | 2175   | AX833435 | AX833435 Sequence  |
| 19         | 858.4  | 36.3        | 2175   | AK095160 | AK095160 Homo sapi |

|    |       |      |        |    |          |
|----|-------|------|--------|----|----------|
| 20 | 806.8 | 34.1 | 169891 | 9  | AC020907 |
| 21 | 779.6 | 33.0 | 1580   | 10 | BC072688 |
| 22 | 600   | 25.4 | 600    | 6  | AX530376 |
| 23 | 391.2 | 16.6 | 614    | 6  | AR263893 |
| 24 | 380.2 | 16.1 | 497    | 6  | AX698852 |
| 25 | 325   | 13.8 | 21784  | 6  | AR255886 |
| 26 | 325   | 13.6 | 174875 | 2  | AC024682 |
| 27 | 321.8 | 13.6 | 2547   | 6  | AX747294 |
| 28 | 321.8 | 13.6 | 2547   | 9  | AK091988 |
| 29 | 297   | 12.6 | 297    | 11 | G06603   |
| 30 | 249.8 | 10.6 | 266    | 6  | AX335256 |
| 31 | 249.8 | 10.6 | 266    | 6  | AX409165 |
| 32 | 235   | 9.9  | 2831   | 6  | CO842731 |
| 33 | 235   | 9.9  | 2831   | 9  | AK125670 |
| 34 | 192   | 8.1  | 601    | 6  | AR255897 |
| 35 | 191   | 8.1  | 601    | 6  | AR255898 |
| 36 | 170   | 7.2  | 133276 | 2  | AC087147 |
| 37 | 170   | 7.2  | 216120 | 2  | AC021477 |
| 38 | 169.4 | 7.2  | 601    | 6  | AR255896 |
| 39 | 168.4 | 7.1  | 239049 | 2  | AC111870 |
| 40 | 168.4 | 7.1  | 244224 | 2  | AC125992 |
| 41 | 163.8 | 6.9  | 177    | 1  | G14606   |
| 42 | 150.8 | 6.4  | 2237   | 9  | AY190317 |
| 43 | 148.4 | 6.3  | 1078   | 6  | AX190545 |
| 44 | 148.4 | 6.3  | 1314   | 6  | AX190543 |
| 45 | 148.4 | 6.3  | 1314   | 6  | AX190546 |

## ALIGNMENTS

|          |             |
|----------|-------------|
| AC020907 | Homo sapi   |
| BC072688 | Rattus no   |
| AX530376 | Sequence    |
| AR263893 | Sequence    |
| AX698852 | Sequence    |
| AR255886 | Sequence    |
| AC024682 | Homo sapi   |
| AX747294 | Sequence    |
| AK091988 | Homo sapi   |
| G06603   | human STS W |
| AX335256 | Sequence    |
| AX409165 | Sequence    |
| CO842731 | Sequence    |
| AK125670 | Homo sapi   |
| AR255897 | Sequence    |
| AR255898 | Sequence    |
| AC087147 | Mus muscu   |
| AC021477 | Mus muscu   |
| AR255896 | Sequence    |
| AC111870 | Rattus no   |
| AC125992 | Rattus no   |
| G14606   | human STS S |
| AY190317 | Homo sapi   |
| AX190545 | Sequence    |
| AX190543 | Sequence    |
| AX190546 | Sequence    |

|                            |  |   |     |        |                 |
|----------------------------|--|---|-----|--------|-----------------|
| RESULT 1                   | AR220138                                   | Sequence 3 from patent US-6423543.                        | DNA | Linear | PAT 26-SEP-2002 |
| LOCUS                      | AR220138                                   |   |     |        |                 |
| DEFINITION                 | Sequence 3 from patent US-6423543.         |   |     |        |                 |
| ACCESSION                  | AR220138                                   |   |     |        |                 |
| VERSION                    | AR220138.1                                 | GI:23374581   |     |        |                 |
| KEYWORDS                   | Unknown.                                   |   |     |        |                 |
| SOURCE                     | Unknown.                                   |   |     |        |                 |
| ORGANISM                   | Unclassified.                              |   |     |        |                 |
| REFERENCE                  | 1 (bases 1 to 2363)                        |   |     |        |                 |
| AUTHORS                    | Marcotte, P.A. and Cowert, L.M.            |   |     |        |                 |
| TITLE                      | Antisense modulation of heparin expression |   |     |        |                 |
| JOURNAL                    | Patent: US-6423543-A 3 23-JUL-2002;        |   |     |        |                 |
| FEATURES                   | Location/Qualifiers                        |   |     |        |                 |
| source                     | 1..2363                                    |   |     |        |                 |
|                            | /organism="unknown"                        |   |     |        |                 |
|                            | /mol_type="genomic DNA"                    |   |     |        |                 |
| ORIGIN                     |  |   |     |        |                 |
| Query Match                | 100.0%; Score 2363; DB 6; Length 2363;     |   |     |        |                 |
| Best Local Similarity      | 100.0%; Pred. No. 0;                       |   |     |        |                 |
| Matches 2363; Conservative | 0; Mismatches 0; Indels 0; Gaps 0;         |   |     |        |                 |
| QY                         | 1  | TCGAGCCGCTTTCCAGGAGCCTTACTGAGGCGCCACAGGTAGGAGGCTGAGCTTAC  |     |        |                 |
| DB                         | 1  | TCGAGCCGCTTTCCAGGAGCCTTACTGAGGCGCCACAGGTAGGAGGCTGAGCTTAC  |     |        |                 |
| QY                         | 61   | AGGCCCCAGCCAGCCGCTTGCCTCCAGGCGCCGCTGCTCGGGGACCAATGCTCC    |     |        |                 |
| DB                         | 61   | AGGCCCCAGCCAGCCGCTTGCCTCCAGGCGCCGCTGCTCGGGGACCAATGCTCC    |     |        |                 |
| QY                         | 121  | TGCGCAGGCTTGAGAGTACGAGCCGAGCCTTACTGAGGCTTCCGCTTACGAGCTTAC |     |        |                 |
| DB                         | 121  | TGCGCAGGCTTGAGAGTACGAGCCGAGCCTTACTGAGGCTTCCGCTTACGAGCTTAC |     |        |                 |
| QY                         | 181  | TGAGCCCGAGGTTAAGAGCAGGCGCCGAGTACAGTTCAGGCTTACGAGCAGGG     |     |        |                 |
| DB                         | 181  | TGAGCCCGAGGTTAAGAGCAGGCGCCGAGTACAGTTCAGGCTTACGAGCAGGG     |     |        |                 |
| QY                         | 241  | TTCCCTCATCCCGCCAGCCGCTTATGCGCCTTATGAGGGGCTTCTGGGAGC       |     |        |                 |
| DB                         | 241  | TTCCCTCATCCCGCCAGCCGCTTATGCGCCTTATGAGGGGCTTCTGGGAGC       |     |        |                 |

Db 241 TTCCCTCAATCCCCCAACCCAGCTTAATGCCCCACCTCTTAATAGAGGGGTTCTGGAGAC 300  
Qy 301 TGAAGAGGGGACATAGAGCTCTCCCAAGCACTAGGTGTTGTCCTGCTCTTCTT 360  
Db 301 TGAAGAGGGGACATAGAGCTCTCCCAAGCACTAGGTGTTGTCCTGCTCTTCTT 360  
Qy 361 CAGACTCAGCGGTGAGACCCAGTCTCTTCTCCAGACCCAGAGTTTCAGCCCTAG 420  
Db 361 CAGACTCAGCGGTGAGACCCAGTCTCTTCTCCAGACCCAGAGTTTCAGCCCTAG 420  
Qy 421 GCCCTCTCTCCTCATACTAGAGAGTCTGAGCCCAATTCCTCTCTTCTCCAGACTTA 480  
Db 421 GCCCTCTCTCCTCATACTAGAGAGTCTGAGCCCAATTCCTCTCTTCTCCAGACTTA 480  
Qy 481 TGAATTTAGGTCTCAGCTGTCTCTCTCTCAAAACCGGAGATCTCAGTCCCTCTCCAC 540  
Db 481 TGAATTTAGGTCTCAGCTGTCTCTCTCTCAAAACCGGAGATCTCAGTCCCTCTCCAC 540  
Qy 541 CAGGCTCAGGATGGGGGTTCCCATCCCTGCAAAATCCAGAGGTCCCGGCTGCTGTC 600  
Db 541 CAGGCTCAGGATGGGGGTTCCCATCCCTGCAAAATCCAGAGGTCCCGGCTGCTGTC 600  
Qy 601 GACA CTGACCCCATCTTTGAA CCGAGCCCAATCTGCGTCTGATCA CAGCGTCTG 660  
Db 601 GACA CTGACCCCATCTTTGAA CCGAGCCCAATCTGCGTCTGATCA CAGCGTCTG 660  
Qy 661 CCAAGGCCCAATCCCTTACAGCTGTGCTGAGATGAGACGCTGGAGATCTGGGGGGCCAGAGT 720  
Db 661 CCAAGGCCCAATCCCTTACAGCTGTGCTGAGATGAGACGCTGGAGATCTGGGGGGCCAGAGT 720  
Qy 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTCCATCTCTCTCAAGTCCACCC 780  
Db 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTCCATCTCTCTCAAGTCCACCC 780  
Qy 781 TGGCCCAAGAGGTCAAGCCAGGGAATCATTTAACAGAGGCACTGACATGACCAAGAGAG 840  
Db 781 TGGCCCAAGAGGTCAAGCCAGGGAATCATTTAACAGAGGCACTGACATGACCAAGAGAG 840  
Qy 841 GGTGGCCGAGACTGTGCAATGCTGTCCAGACCCAGAGGTGAGAGTCTCACTGGGGAGCC 900  
Db 841 GGTGGCCGAGACTGTGCAATGCTGTCCAGACCCAGAGGTGAGAGTCTCACTGGGGAGCC 900  
Qy 901 CTGCTACTTCTGACAGCCATCGGGGCGGATCTGAGCCATTGTTGCTCTCTAG 960  
Db 901 CTGCTACTTCTGACAGCCATCGGGGCGGATCTGAGCCATTGTTGCTCTCTAG 960  
Qy 961 AGTACCAAGAGCGCTGTGTACCAATGCAAGGTCAAGCTCTGCGAGCTCGGCTCATGCTC 1020  
Db 961 AGTACCAAGAGCGCTGTGTACCAATGCAAGGTCAAGCTCTGCGAGCTCGGCTCATGCTC 1020  
Qy 1021 TTTTACCAAGAGCGGAGGAGCGTGTGCGTGTGCTCTGCGCTCCAAAGCCAGAGGTA 1080  
Db 1021 TTTTACCAAGAGCGGAGGAGCGTGTGCGTGTGCTCTGCGCTCCAAAGCCAGAGGTA 1080  
Qy 1081 GCCGAGCTCAGCTCGAGAGATGAGGCTTCTCAAGGCGCATGACCACTCCGACTGGAC 1140  
Db 1081 GCCGAGCTCAGCTCGAGAGATGAGGCTTCTCAAGGCGCATGACCACTCCGACTGGAC 1140  
Qy 1141 GTGCGAAACGCGGAGCGCCATGAGCAAGTGGGCTTTCTTGTGTGAGAGAGGAGCTG 1200  
Db 1141 GTGCGAAACGCGGAGCGCCATGAGCAAGTGGGCTTTCTTGTGTGAGAGAGGAGCTG 1200  
Qy 1201 CCCCACACCCAGAGGCTGTGAGAGTCACTCCGTTGTGTATTTCCCAAGAGGCGTTTC 1260  
Db 1201 CCCCACACCCAGAGGCTGTGAGAGTCACTCCGTTGTGTATTTCCCAAGAGGCGTTTC 1260  
Qy 1261 TTGGCCCGCATCTGCAAGAGCTGTGAGCGAGAGGCTGCCGTTGAGACCGGATGCTGGGA 1320  
Db 1261 TTGGCCCGCATCTGCAAGAGCTGTGAGCGAGAGGCTGCCGTTGAGACCGGATGCTGGGA 1320  
Qy 1321 GGGCGGAGCACAGCTTGGGCGGTGGCGGTGGCAAGTCAAGCTTTGCTATGATGAGACA 1380  
Db 1321 GGGCGGAGCACAGCTTGGGCGGTGGCGGTGGCAAGTCAAGCTTTGCTATGATGAGACA 1380

Qy 1381 CACCTGTGGGGGATCCCTGCTCTCCGGGAGCTGGGTGCTGACAGCGCCACTGCTTC 1440  
Db 1381 CACCTGTGGGGGATCCCTGCTCTCCGGGAGCTGGGTGCTGACAGCGCCACTGCTTC 1440  
Qy 1441 CCGAGGCGGAACCGGGTCTGTCTCCAGTGGCGAGTGTTCCTGGTGGCCAGAGCC 1500  
Db 1441 CCGAGGCGGAACCGGGTCTGTCTCCAGTGGCGAGTGTTCCTGGTGGCCAGAGCC 1500  
Qy 1501 TCTCCCAAGGCTGTGACGCTGGGGGTGAGGCTGTGTCTTACCAACGGGGGCTATCTTCC 1560  
Db 1501 TCTCCCAAGGCTGTGACGCTGGGGGTGAGGCTGTGTCTTACCAACGGGGGCTATCTTCC 1560  
Qy 1561 TTTTGGAGACCCCAACAGCGAGAGAGAGAGAGATTTGCTCCCTGCTGCACTCTCCAGT 1620  
Db 1561 TTTTGGAGACCCCAACAGCGAGAGAGAGAGATTTGCTCCCTGCTGCACTCTCCAGT 1620  
Qy 1621 CCCCTGCTCTCAAGATAATTCAGCTGTGTGCTCTCCAGCTGCGGCGAGGCTTG 1680  
Db 1621 CCCCTGCTCTCAAGATAATTCAGCTGTGTGCTCTCCAGCTGCGGCGAGGCTTG 1680  
Qy 1681 GTGATGCGAAGATCTGTACCGTGAAGGCTGGGGCAACAGATCTATGCGCAAG 1740  
Db 1681 GTGATGCGAAGATCTGTACCGTGAAGGCTGGGGCAACAGATCTATGCGCAAG 1740  
Qy 1741 GCGGGGGTACTCCAGAGAGGCTCGAGTCCCATATACAGATATATGTCTCAATGGCGCT 1800  
Db 1741 GCGGGGGTACTCCAGAGAGGCTCGAGTCCCATATACAGATATATGTCTCAATGGCGCT 1800  
Qy 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGAGTGC 1860  
Db 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGAGTGC 1860  
Qy 1861 ATTGATGCTGCGAGGCGGACAGCGGTGTCTCTTGTGTGAGAGACAGATCTCTCGG 1920  
Db 1861 ATTGATGCTGCGAGGCGGACAGCGGTGTCTCTTGTGTGAGAGACAGATCTCTCGG 1920  
Qy 1921 ACGCAGCTGTGGCGGCTGTGAGCATGTGAGTTGGGAGCACTGGCTGTGCCCTGGCCAG 1980  
Db 1921 ACGCAGCTGTGGCGGCTGTGAGCATGTGAGTTGGGAGCACTGGCTGTGCCCTGGCCAG 1980  
Qy 1981 AAGCAGAGCGTCTACCAAGATCAAGTCTTCGAGAGTGAATCTTCAAGGCCATTAAG 2040  
Db 1981 AAGCAGAGCGTCTACCAAGATCAAGTCTTCGAGAGTGAATCTTCAAGGCCATTAAG 2040  
Qy 2041 ACTCATCTCGAAGCCAGCGGATGTGAACCAAGTCTGACCGGTGTCTCTCGTGGCA 2100  
Db 2041 ACTCATCTCGAAGCCAGCGGATGTGAACCAAGTCTGACCGGTGTCTCTCGTGGCA 2100  
Qy 2101 GCTTCAGAGGCGGAGGTGATCCCGGTGTGAGATCAACGCTGGGCGAGATGGAGCT 2160  
Db 2101 GCTTCAGAGGCGGAGGTGATCCCGGTGTGAGATCAACGCTGGGCGAGATGGAGCT 2160  
Qy 2161 TTTTCTTCTTGGGCGCGGTCAAGGTCAGGATCCCTCTCAAGGATCTCTCTTC 2220  
Db 2161 TTTTCTTCTTGGGCGCGGTCAAGGTCAGGATCCCTCTCAAGGATCTCTCTTC 2220  
Qy 2221 CACAGTGGCGGCGGCTCACTGAGCCCGAGACCAACCTCAACCTCTGACCCCAATGT 2280  
Db 2221 CACAGTGGCGGCGGCTCACTGAGCCCGAGACCAACCTCAACCTCTGACCCCAATGT 2280  
Qy 2281 AAATATTTGTTCTGCTGTCTGAGGACTCTGTCTAGGTGCTCCCTGATGATGAGATCTCTTT 2340  
Db 2281 AAATATTTGTTCTGCTGTCTGAGGACTCTGTCTAGGTGCTCCCTGATGATGAGATCTCTTT 2340  
Qy 2341 AAATATTAAGATGTTTGAAT 2363  
Db 2341 AAATATTAAGATGTTTGAAT 2363

RESULT 2  
AX336397  
LOCUS AX336397 2363 bp DNA linear PAT 09-JAN-2002



DEFINITION Sequence 6906 from Patent WO0194629.  
 ACCESSION AX336397  
 VERSION AX336397.1 GI:18127116  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, C.,  
 Horrigan, S., Soppel, D.R. and Weaver, Z.  
 TITLE Cancer gene determination and therapeutic screening using signature  
 gene sets  
 JOURNAL Patent: WO 0194629-A 6906 13-DEC-2001;  
 Avalon Pharmaceuticals (US)  
 FEATURES  
 Source 1..2363  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Query Match 100.0%; Score 2363; DB 6; Length 2363;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

721 GGGCTGGGCTGGGGCTCCCCAGAGGCTGCTCCCCCTCAATCTCTCAAGTCCACCC 780  
 721 GGGCTGGGCTGGGGCTCCCCAGAGGCTGCTCCCCCTCAATCTCTCAAGTCCACCC 780  
 781 TGGCCCAAGAGTGCAGCCAGGAATCAATTAACAAGAGAGTGCATGCGCAGAAAGAG 840  
 781 TGGCCCAAGAGTGCAGCCAGGAATCAATTAACAAGAGAGTGCATGCGCAGAAAGAG 840  
 841 GGTGGCCGGAATGTGCAATGTCTGCTCAAGACCAAGGTGGAGCTTCACTGCGGAGAC 900  
 841 GGTGGCCGGAATGTGCAATGTCTGCTCAAGACCAAGGTGGAGCTTCACTGCGGAGAC 900  
 841 GGTGGCCGGAATGTGCAATGTCTGCTCAAGACCAAGGTGGAGCTTCACTGCGGAGAC 900  
 901 CTGCTACTTTCAGAGGCAATTCGGGCGGCAATCTGGGCGCATTTGGCTGTCTCTCAG 960  
 901 CTGCTACTTTCAGAGGCAATTCGGGCGGCAATCTGGGCGCATTTGGCTGTCTCTCAG 960  
 901 CTGCTACTTTCAGAGGCAATTCGGGCGGCAATCTGGGCGCATTTGGCTGTCTCTCAG 960  
 961 AGTACCAGAGAGCCGCTGTATCCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
 961 AGTACCAGAGAGCCGCTGTATCCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
 1021 TTTGACAAAGACGGAAGGAGAGTGGCGGCTGTGCTCTTCGCGCTCAACGCGAGGTA 1080  
 1021 TTTGACAAAGACGGAAGGAGAGTGGCGGCTGTGCTCTTCGCGCTCAACGCGAGGTA 1080  
 1081 GCGGAGCTCAAGCTGCGGAGAGATGGGCTTCTCTCAAGGCACTGACCACTCGAGCTGAC 1140  
 1081 GCGGAGCTCAAGCTGCGGAGAGATGGGCTTCTCTCAAGGCACTGACCACTCGAGCTGAC 1140  
 1141 GTGCAACAGGCGGCGGCGCAATGGCAAGTGGGCTTCTCTGTGTGAGAGAGGAGGCTG 1200  
 1141 GTGCAACAGGCGGCGGCGCAATGGCAAGTGGGCTTCTCTGTGTGAGAGAGGAGGCTG 1200  
 1201 CCCCACACCAAGAGGCTGTGAGAGTATCTCCGTGTGTATTTGCCCAAGAGGCGGTTTC 1260  
 1201 CCCCACACCAAGAGGCTGTGAGAGTATCTCCGTGTGTATTTGCCCAAGAGGCGGTTTC 1260  
 1201 CCCCACACCAAGAGGCTGTGAGAGTATCTCCGTGTGTATTTGCCCAAGAGGCGGTTTC 1260  
 1261 TTGGCGGCTATCTGCCAAGACTGTGGCGGAGAGAGTGGCGGCGGCGGCGGCGGCGG 1320  
 1261 TTGGCGGCTATCTGCCAAGACTGTGGCGGAGAGAGTGGCGGCGGCGGCGGCGGCGG 1320  
 1321 GGGCGGAGACCAAGCTTGGGCGGCTGGCGGAGTGAAGCTTGGCTATGATGAGACA 1380  
 1321 GGGCGGAGACCAAGCTTGGGCGGCTGGCGGAGTGAAGCTTGGCTATGATGAGACA 1380  
 1381 CACTCTGTGGGAGATCCCTGCTCTCCGCGGAGCTGGGCTGAGACGCGGCACTGCTTC 1440  
 1381 CACTCTGTGGGAGATCCCTGCTCTCCGCGGAGCTGGGCTGAGACGCGGCACTGCTTC 1440  
 1441 CCGAGCGGAACCGGCTCTGTCCGATGGCGAGATTTTCCCGGTGCGTGGCCAGGCC 1500  
 1441 CCGAGCGGAACCGGCTCTGTCCGATGGCGAGATTTTCCCGGTGCGTGGCCAGGCC 1500  
 1501 TCTCCCAAGGCTGCAAGCTGGGAGTGCAGGCTGTGTCTAACAAGGAGGCTATCTTCCC 1560  
 1501 TCTCCCAAGGCTGCAAGCTGGGAGTGCAGGCTGTGTCTAACAAGGAGGCTATCTTCCC 1560  
 1561 TTTCCGAGACCCCAACAGAGAGAGAAACAGCAAGATTTTCCCGGTGCGTGGCCAGGCC 1620  
 1561 TTTCCGAGACCCCAACAGAGAGAGAAACAGCAAGATTTTCCCGGTGCGTGGCCAGGCC 1620  
 1621 TTTCCGAGACCCCAACAGAGAGAGAAACAGCAAGATTTTCCCGGTGCGTGGCCAGGCC 1680  
 1621 TTTCCGAGACCCCAACAGAGAGAGAAACAGCAAGATTTTCCCGGTGCGTGGCCAGGCC 1680  
 1681 GTGGATGGCAAGATCTGATCCGTAAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGG 1740  
 1681 GTGGATGGCAAGATCTGATCCGTAAGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGG 1740  
 1741 GCGGAGGTAATCCAGAGAGCTGAGTCCCATATATGACGAATGATGTCTGCAATGGCGCT 1800  
 1741 GCGGAGGTAATCCAGAGAGCTGAGTCCCATATATGACGAATGATGTCTGCAATGGCGCT 1800  
 1741 GCGGAGGTAATCCAGAGAGCTGAGTCCCATATATGACGAATGATGTCTGCAATGGCGCT 1800

QY 1801 GACTTCTATGGAACACAGATCAAGCCCAAGATGTTCTGTGTGCTACACCCGAGGGTGGC 1860  
|  
|  
|  
Db 1801 GACTTCTATGGAACACAGATCAAGCCCAAGATGTTCTGTGTGCTACACCCGAGGGTGGC 1860  
QY 1861 ATTGATGCTGTGCGACAGGGGCGACAGCGGTGTGCTCTTTGTGTGTGAGAACAGCATCTTCGG 1920  
|  
|  
|  
Db 1861 ATTGATGCTGTGCGACAGGGGCGACAGCGGTGTGCTCTTTGTGTGTGAGAACAGCATCTTCGG 1920  
QY 1921 ACGCAGAGTGGGGGCTGTGTGTGAGATTGTGTGTGGGGCACTGTGTGGCTGGGCCAG 1980  
|  
|  
|  
Db 1921 ACGCAGAGTGGGGGCTGTGTGTGAGATTGTGTGTGGGGCACTGTGTGGCTGGGCCAG 1980  
QY 1981 AAGCAGAGGCTGTCAACCAAAAGTCAAGTACTTCCGGAGTGAATCTTCCAGGCCATAAG 2040  
|  
|  
|  
Db 1981 AAGCAGAGGCTGTCAACCAAAAGTCAAGTACTTCCGGAGTGAATCTTCCAGGCCATAAG 2040  
QY 2041 ACTCATCTCCGAAGCCAGCGGCGATGTGACCCAGCTCTGAACCGGTGTGCTTCTGCTGGCA 2100  
|  
|  
|  
Db 2041 ACTCATCTCCGAAGCCAGCGGCGATGTGACCCAGCTCTGAACCGGTGTGCTTCTGCTGGCA 2100  
QY 2101 GCCCTCAAGGGCCGAGGGTATCCGGTGTGTGGATCCAGCTGGGGCGAGGATGGAGCT 2160  
|  
|  
|  
Db 2101 GCCCTCAAGGGCCGAGGGTATCCGGTGTGTGGATCCAGCTGGGGCGAGGATGGAGCT 2160  
QY 2161 TTTTCTTCTTGGGCGCGGTCCAGAGTTCAGAGACACCCCTCTCCAGAGTCTCTCTTC 2220  
|  
|  
|  
Db 2161 TTTTCTTCTTGGGCGCGGTCCAGAGTTCAGAGACACCCCTCTCCAGAGTCTCTCTTC 2220  
QY 2221 CACAGTGGCGGGCCCACTCAGCCCCGAGAGCCACCACTCAACCTCTCTCTGACCCCAATGT 2280  
|  
|  
|  
Db 2221 CACAGTGGCGGGCCCACTCAGCCCCGAGAGCCACCACTCAACCTCTCTCTGACCCCAATGT 2280  
QY 2281 AAATATGTTCTGTCTGTCTGTGGAGTCTCTGTGTAGTGTGCTGTGTGTGTGTGTGTGT 2340  
|  
|  
|  
Db 2281 AAATATGTTCTGTCTGTCTGTGGAGTCTCTGTGTAGTGTGCTGTGTGTGTGTGTGTGT 2340  
QY 2341 AAATATTAAGATGTTTGTGATT 2363  
|  
|  
|  
Db 2341 AAATATTAAGATGTTTGTGATT 2363

RESULT 3  
AX411046 2363 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 3693 from Patent WO0229103.  
DEFINITION AX411046  
ACCESSION AX411046  
VERSION AX411046.1 GI:21443751  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1  
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
Gene expression profiles in liver cancer  
Patent: WO 0229103-A 3693 11-APR-2002;  
JOURNAL GENE LOGIC INC (US)  
FEATURES  
Location/Qualifiers  
1..2363  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. X07732"

Origin  
Query Match 100.0%; Score 2363; DB 6; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 AGGCCCAAGCCCAACCGCTGTGCTCCAGAGCCGCGCTGTGCGGGCCACCATGTCTC 120  
|  
|  
|  
Db 61 AGGCCCAAGCCCAACCGCTGTGCTCCAGAGCCGCGCTGTGCGGGCCACCATGTCTC 120  
QY 121 TGGCAGGCTGTGAAGATTGACCCGAGCCCGGACTATCTGTAGAGCTTCGCCCACTTGC 180  
|  
|  
|  
Db 121 TGGCAGGCTGTGAAGATTGACCCGAGCCCGGACTATCTGTAGAGCTTCGCCCACTTGC 180  
QY 181 TGGACCCCAAGGGTAAGGACAAAGGGCCCGGAGCTCAAGTTCAGAGCCCTGAGAGAG 240  
|  
|  
|  
Db 181 TGGACCCCAAGGGTAAGGACAAAGGGCCCGGAGCTCAAGTTCAGAGCCCTGAGAGAG 240  
QY 241 TTCCCTATCCCCCAACCCAGCTTAATGCCCACCTCTTAATAGAGGGTTCTTGGAGAC 300  
|  
|  
|  
Db 241 TTCCCTATCCCCCAACCCAGCTTAATGCCCACCTCTTAATAGAGGGTTCTTGGAGAC 300  
QY 301 TGAAGAGGGGGCACTATGAGTCTCCCAAGACCTTAGTGTGTGTGTGTGTGTGTGT 360  
|  
|  
|  
Db 301 TGAAGAGGGGGCACTATGAGTCTCCCAAGACCTTAGTGTGTGTGTGTGTGTGTGT 360  
QY 361 CAGACTAGCGGTGTGACCCAGCTTCTCTCCAGAGCCAGAGATTCCAGGCTCTCAG 420  
|  
|  
|  
Db 361 CAGACTAGCGGTGTGACCCAGCTTCTCTCCAGAGCCAGAGATTCCAGGCTCTCAG 420  
QY 421 GCCCTCTCTCTCATATCTAGAGAGTCTGCGCCCAATTCCTCTTCCAGAGCTTA 480  
|  
|  
|  
Db 421 GCCCTCTCTCTCATATCTAGAGAGTCTGCGCCCAATTCCTCTTCCAGAGCTTA 480  
QY 481 TGATTTAGGTCTCTAGCTGTCTCTCTCTCAAAAGGGATCTCAATCTCCCTGTCTCAC 540  
|  
|  
|  
Db 481 TGATTTAGGTCTCTAGCTGTCTCTCTCTCAAAAGGGATCTCAATCTCCCTGTCTCAC 540  
QY 541 CAGGCTCAGGCAATGAGGGGCTCCCATCCCTGCAATCCAGGAGTCCCGCTGTGTGCA 600  
|  
|  
|  
Db 541 CAGGCTCAGGCAATGAGGGGCTCCCATCCCTGCAATCCAGGAGTCCCGCTGTGTGCA 600  
QY 541 CAGGCTCAGGCAATGAGGGGCTCCCATCCCTGCAATCCAGGAGTCCCGCTGTGTGCA 600  
|  
|  
|  
Db 541 CAGGCTCAGGCAATGAGGGGCTCCCATCCCTGCAATCCAGGAGTCCCGCTGTGTGCA 600  
QY 601 GACACTGACCCCATCTCTTGAACCCAGCCCAATCTGCTGCTGTATCAAGCGGTGTCTGG 660  
|  
|  
|  
Db 601 GACACTGACCCCATCTCTTGAACCCAGCCCAATCTGCTGCTGTATCAAGCGGTGTCTGG 660  
QY 661 CCAAGGCCAGTCTCTTGAACCCAGCCCAATCTGCTGCTGTATCAAGCGGTGTCTGG 720  
|  
|  
|  
Db 661 CCAAGGCCAGTCTCTTGAACCCAGCCCAATCTGCTGCTGTATCAAGCGGTGTCTGG 720  
QY 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGGCAATCTCTCAGAGTCTCCAGCC 780  
|  
|  
|  
Db 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGGCAATCTCTCAGAGTCTCCAGCC 780  
QY 781 TGGCCAGAGGTCAGCCAGGAGATCTTAACAAGAGAGTGAATGAGCGCGAGAGAG 840  
|  
|  
|  
Db 781 TGGCCAGAGGTCAGCCAGGAGATCTTAACAAGAGAGTGAATGAGCGCGAGAGAG 840  
QY 841 GGTGGCCGAGTGTGCTGCTGTCTGCAAGCCAGGTGGAGTCTCACTGTGGGGACC 900  
|  
|  
|  
Db 841 GGTGGCCGAGTGTGCTGCTGTCTGCAAGCCAGGTGGAGTCTCACTGTGGGGACC 900  
QY 901 CTGCTACTTGTGAACAGCATCGGGGGGAGTCTGTGGGCAATTGAGCTGTCTCTCAGG 960  
|  
|  
|  
Db 901 CTGCTACTTGTGAACAGCATCGGGGGGAGTCTGTGGGCAATTGAGCTGTCTCTCAGG 960  
QY 961 AGTACACAGAGCGCTGTATCCAGTGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
|  
|  
|  
Db 961 AGTACACAGAGCGCTGTATCCAGTGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
QY 1021 TTTGACAAAGAGGAGAGAGTGTGCGGCTGTGCTCTGCGGCTCAACGCCAGGGTA 1080  
|  
|  
|  
Db 1021 TTTGACAAAGAGGAGAGAGTGTGCGGCTGTGCTCTGCGGCTCAACGCCAGGGTA 1080  
QY 1081 GCCGAGCTCACTGTGAGAGAGTGTGCTCTGAGGCACTGACCCCACTCGAGCTGAC 1140  
|  
|  
|  
Db 1081 GCCGAGCTCACTGTGAGAGAGTGTGCTCTGAGGCACTGACCCCACTCGAGCTGAC 1140  
QY 1141 GTGCGAAAGCGCGGGCGCAATGAGTGTGCGGCTTCTTGTGTGTGAGAGAGGAGCTG 1200

|    |      |  |      |
|----|------|--|------|
| Db | 1141 | GTGCGAATGGCGGGGCCAATGGCACTCGGGCTTCTCTGTGTGATGACGAGGGAGGCTG   | 1200 |
| Oy | 1201 | CCCCACATCCAGAGGCTGTGAGAGTCACTTCCTGTGTGATGATGCCAGAGGCGCTTC    | 1260 |
| Db | 1201 | CCCCACATCCAGAGGCTGTGAGAGTCACTTCCTGTGTGATGATGCCAGAGGCGCTTC    | 1260 |
| Oy | 1261 | TTTGGCGGCACTCTGCCAAGACTGTGTGGCCGACAGAAAGCTGCCGTGTGACATGTGTGG | 1320 |
| Db | 1261 | TTTGGCGGCACTCTGCCAAGACTGTGTGGCCGACAGAAAGCTGCCGTGTGACATGTGTGG | 1320 |
| Oy | 1321 | GAGCGGGACACACAGCTTGAGGCGAGTGGCGAGCAATGACGCTTCGCTATGATGAGACA  | 1380 |
| Db | 1321 | GAGCGGGACACACAGCTTGAGGCGAGTGGCGAGCAATGACGCTTCGCTATGATGAGACA  | 1380 |
| Oy | 1381 | CACCTCTGTGGGGGATCCCTGTCTCCGGGAGCTGGGTGTGACAGCGGCCCATCTGCTTC  | 1440 |
| Db | 1381 | CACCTCTGTGGGGGATCCCTGTCTCCGGGAGCTGGGTGTGACAGCGGCCCATCTGCTTC  | 1440 |
| Oy | 1441 | CCGAGCGGAAACCGGGTCTGTCCCGATGGCGAGTGTTCGCGGTGCCGTGGCCAGGC     | 1500 |
| Db | 1441 | CCGAGCGGAAACCGGGTCTGTCCCGATGGCGAGTGTTCGCGGTGCCGTGGCCAGGC     | 1500 |
| Oy | 1501 | TCTCCCAACGGCTGTGACGTGGGGGTGACGCTGTGTCAACGAGGGGGCTATCTTCC     | 1560 |
| Db | 1501 | TCTCCCAACGGCTGTGACGTGGGGGTGACGCTGTGTGTCAACGAGGGGGCTATCTTCC   | 1560 |
| Oy | 1561 | TTTGGGAGCCCCAACAGGAGAGAACAGCAACGATATTGCTGTATCCATCTCCAGT      | 1620 |
| Db | 1561 | TTTGGGAGCCCCAACAGGAGAGAACAGCAACGATATTGCTGTATCCATCTCCAGT      | 1620 |
| Oy | 1621 | CCCGTCCCTTCACAGAAATACATCCAGCTGTGTGTCTCCAGCTGCCGCGCAGGCTTG    | 1680 |
| Db | 1621 | CCCGTCCCTTCACAGAAATACATCCAGCTGTGTGTCTCCAGCTGCCGCGCAGGCTTG    | 1680 |
| Oy | 1681 | GTGATGGCAAGATCTGTAACTGGTACGGGCTGGGGCAACAGCGATCTATATGGCAACAG  | 1740 |
| Db | 1681 | GTGATGGCAAGATCTGTAACTGGTACGGGCTGGGGCAACAGCGATCTATATGGCAACAG  | 1740 |
| Oy | 1741 | GCGGGGATCTCCAGAGAGGCTCGAATCCCAATAATCAGCAATGATGTCTGCAATGGGCT  | 1800 |
| Db | 1741 | GCGGGGATCTCCAGAGAGGCTCGAATCCCAATAATCAGCAATGATGTCTGCAATGGGCT  | 1800 |
| Oy | 1801 | GACTTCTATGAAAAACAGATCAAGGCCAAGATGTTCTGTGTGTGCTACCCCGAGGGTGGC | 1860 |
| Db | 1801 | GACTTCTATGAAAAACAGATCAAGGCCAAGATGTTCTGTGTGTGCTACCCCGAGGGTGGC | 1860 |
| Oy | 1861 | ATTGATGCTGTCCAGGCGGACAGGCGGTGTCTCTTGTGTGTGAGACAGCATCTTCGG    | 1920 |
| Db | 1861 | ATTGATGCTGTCCAGGCGGACAGGCGGTGTCTCTTGTGTGTGAGACAGCATCTTCGG    | 1920 |
| Oy | 1921 | ACGCGACGTTGAGCGGCTGTGTGGCATTTGTGATTTGGGGCATTTGGCTGTGCTTGCCAG | 1980 |
| Db | 1921 | ACGCGACGTTGAGCGGCTGTGTGGCATTTGTGATTTGGGGCATTTGGCTGTGCTTGCCAG | 1980 |
| Oy | 1981 | AAGCCAGGCGTCTACACCAAAATCAGTACTTCCGAGAGTGTGATCTTCCAGGCCATAAG  | 2040 |
| Db | 1981 | AAGCCAGGCGTCTACACCAAAATCAGTACTTCCGAGAGTGTGATCTTCCAGGCCATAAG  | 2040 |
| Oy | 2041 | ACTCACTCCGAAGCCAGGCGGATGTGTACCCAGCTCTGACCGGTGGCTTCTCGTGGCA   | 2100 |
| Db | 2041 | ACTCACTCCGAAGCCAGGCGGATGTGTACCCAGCTCTGACCGGTGGCTTCTCGTGGCA   | 2100 |
| Oy | 2101 | GCCTTCAGGGCCGAGGTGATCCGCGGTGTGGATCCACGCTGGGCGAGAGTGGAGCT     | 2160 |
| Db | 2101 | GCCTTCAGGGCCGAGGTGATCCGCGGTGTGGATCCACGCTGGGCGAGAGTGGAGCT     | 2160 |
| Oy | 2161 | TTTTCTTCTTGGGCGCGGTCCACAGGTCCAGAGTCCAGGCTCTCTCTTC            | 2220 |
| Db | 2161 | TTTTCTTCTTGGGCGCGGTCCACAGGTCCAGAGTCCAGGCTCTCTCTTC            | 2220 |
| Oy | 2221 | CACAGTGGCGGGCCATCTCAGCGCCGAGACCAACCAACTCACTCTCTGACCCCAATG    | 2280 |
| Db | 2221 | CACAGTGGCGGGCCATCTCAGCGCCGAGACCAACCAACTCACTCTCTGACCCCAATG    | 2280 |

| DB | LOCUS      | SEQUENCE  | ACCESION | VERSION    | KEYWORDS    | SOURCE               | ORGANISM     | REFERENCE  | AUTHORS                                 | TITLE  | JOURNAL |
|----|------------|---|----------|------------|-------------|----------------------|--------------|--|---|--|---------|
| Db | AXS11632/c | Sequence 39 from Patent WO02055705.                       | AXS11632 | AXS11632.1 | GI:23392433 | Homo sapiens (human) | Homo sapiens | Mezes, P.S., Rastelli, L., Herrmann, J.L., MacDougall, J.R., Zhong, H., Casman, S.J., Boldog, F., Shimkete, R.A., Gorman, L., Crasta, O.R., Mysore, K.K., Folkerts, O., Martin, G.B., Eiben, A., Spaderma, S.K., Verne, C.A., Berg, C., Spytek, K.A., DiPippo, V.A., Zehnisen, B.D., Peyman, J.A., Ellerman, K., Stone, D.J., Grose, W.M., Alsbrook, J.P., Lepley, D.M., Rieger, D.K., Burgess, C.E. and Edinger, S. | Proteins and nucleic acid encoding same | Patent: WO 02055705 A 39 18-JUL-2002; Curagen Corporation (US) |         |
| Qy | 2281       | AAATATTGTTCTGCTGTCTGGAGACTCCTGTCTTAGGTGCTGATGATGGATGCTCTT | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2281       | AAATATTGTTCTGCTGTCTGGAGACTCCTGTCTTAGGTGCTGATGATGGATGCTCTT | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2341       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2341       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Db | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 2363     | 2363       | 2363        | 2363                 | 2363         | 2363   | 2363                                    | 2363   | 2363    |
| Qy | 2363       | AAATTAATTAAGATGTTTGAAT                                    | 23       |            |             |                      |              |  |   |  |         |

Db 1943 GCCCTCCCTCCCTCACTACTAGAGAGTCTGGGCCCCCAAAATTCCTCTTTCCCAAGACTTA 1884  
Qy 481 TGAATTTAGAGTCTCAGCTGTCTCTCTCTCCAAACCGGGATCTCAGTCTCCCTGTCCAC 540  
Db 1883 TGAATTTAGAGTCTCAGCTGTCTCTCTCCAAACCGGGATCTCAGTCTCCCTGTCCAC 1824  
Qy 541 CAGGCTCAGAGTGGGGGTCCCAATCCCTGCAAAATCCAGAGGTGCTCCCGCTGTGTGCA 600  
Db 1823 CAGGCTCAGAGTGGGGGTCCCAATCCCTGCAAAATCCAGAGGTGCTCCCGCTGTGTGCA 1764  
Qy 601 GACACTGACCCCATCTTTGAAACCAAGCCCAATCTGCTGTGATCAGAGCTGTCTGG 660  
Db 1763 GACACTGACCCCATCTTTGAAACCAAGCCCAATCTGCTGTGATCAGAGCTGTCTGG 1704  
Qy 661 CCAAGGCCCTAGTCCCTTAGGCTGTGAGTGAAGCCCTGGAGACTGGGGGGCCAGAGCT 720  
Db 1703 CCAAGGCCCTAGTCCCTTAGGCTGTGAGTGAAGCCCTGGAGACTGGGGGGCCAGAGCT 1644  
Qy 721 GGGCTGGGCTGGGCTCCCAAGGCTGCTCCCTGTCATCTCTCAAGAGTCCACCC 780  
Db 1643 GGGCTGGGCTGGGCTCCCAAGGCTGCTCCCTGTCATCTCTCAAGAGTCCACCC 1584  
Qy 781 TGGCCCAAGAGGTGACCCAGGAATCATTAACAAGGCAAGTGAATGCGCAGAGAG 840  
Db 1583 TGGCCCAAGAGGTGACCCAGGAATCATTAACAAGGCAAGTGAATGCGCAGAGAG 1524  
Qy 841 GGTGGCCGGAATGTGCTCCAGACCCAAAGTGGAGTCTCACTGGGGGACC 900  
Db 1523 GGTGGCCGGAATGTGCTCCAGACCCAAAGTGGAGTCTCACTGGGGGACC 1464  
Qy 901 CTGCTACTTTCAGAGCCATCGGGGCGGATCTGTGGCCATTGTGTGCTCTCTCAG 960  
Db 1463 CTGCTACTTTCAGAGCCATCGGGGCGGATCTGTGGCCATTGTGTGCTCTCTCAG 1404  
Qy 961 AGTACCAAGAGCCGCTGTATCCAGTGCAGTCACTCTGCGACGCTGGCTCATGCTC 1020  
Db 1403 AGTACCAAGAGCCGCTGTATCCAGTGCAGTCACTCTGCGACGCTGGCTCATGCTC 1344  
Qy 1021 TTTTACAAGAGAGGAGCGTGGCGGCTGTGTGCTCTGCGCTCCAAAGCCAGAGTA 1080  
Db 1343 TTTTACAAGAGAGGAGCGTGGCGGCTGTGTGCTCTGCGCTCCAAAGCCAGAGTA 1284  
Qy 1081 GCCGAGCTCAGCTCGAGAGATGGGCTTCTCAGGGCATGACCCATCCGAGCTGGAC 1140  
Db 1283 GCCGAGCTCAGCTCGAGAGATGGGCTTCTCAGGGCATGACCCATCCGAGCTGGAC 1224  
Qy 1141 GTGCGAAGCGGCGGCCCATATGCAAGTGGGCTTCTTGTGTGTGAGAGAGGAGCTG 1200  
Db 1223 GTGCGAAGCGGCGGCCCATATGCAAGTGGGCTTCTTGTGTGTGAGAGAGGAGCTG 1164  
Qy 1201 CCCCAACCCAGAGAGGCTGTGAGAGTCACTCCGCTGTGTGATGCCCCAGAGGCGTTTC 1260  
Db 1163 CCCCAACCCAGAGAGGCTGTGAGAGTCACTCCGCTGTGTGATGCCCCAGAGGCGTTTC 1104  
Qy 1261 TTGGCCGCACTCTGCAAGACTGTGGCGCAGAGAGCTGCCGTGAGCCGATCGTGGGA 1320  
Db 1103 TTGGCCGCACTCTGCAAGACTGTGGCGCAGAGAGCTGCCGTGAGCCGATCGTGGGA 1044  
Qy 1321 GGGCGGGAACACAGACTTGGGCGGGTGGCGTGGCAAGTCAAGCTTCCGTATGATGAGCA 1380  
Db 1043 GGGCGGGAACACAGACTTGGGCGGGTGGCGTGGCAAGTCAAGCTTCCGTATGATGAGCA 984  
Qy 1381 CACCTGTGTGGGGATCCCTGTCTCCGGGGAATGGGGTGAAGAGCCGCACTGTCTTC 1440  
Db 983 CACCTGTGTGGGGATCCCTGTCTCCGGGGAATGGGGTGAAGAGCCGCACTGTCTTC 924  
Qy 1441 CCGAGGGAACCGGGTCTGTCCCGATGCGAGTGTGCTGGTGGCGTGGCCAGAGCC 1500  
Db 923 CCGAGGGAACCGGGTCTGTCCCGATGCGAGTGTGCTGGTGGCGTGGCCAGAGCC 864  
Qy 1501 TCTCCCAACGGTCTGCAAGTGGGGGTGCAAGGCTGTGTGCTTACAAGGGGGCTATCTTCC 1560  
Db 863 TCTCCCAACGGTCTGCAAGTGGGGGTGCAAGGCTGTGTGCTTACAAGGGGGCTATCTTCC 804

Qy 1561 TTTGGGGAACCCCAACAGGAGAGAGAGAGAGATATGGCCCTGTCACCTTCCAGT 1620  
Db 803 TTTGGGGAACCCCAACAGGAGAGAGAGAGAGATATATGGCCCTGTCACCTTCCAGT 744  
Qy 1621 CCCCTGCCCCCTCAAGAAATACATCAGCTGTGTGCTCCAGCTGCGGCAAGCCCTG 1680  
Db 743 CCCCTGCCCCCTCAAGAAATACATCAGCTGTGTGCTCCAGCTGCGGCAAGCCCTG 684  
Qy 1681 GTGATGGCAAGATCTGTACCGTGAAGGCTGGGGCAACAGCACTATATGCGCAACAG 1740  
Db 683 GTGATGGCAAGATCTGTACCGTGAAGGCTGGGGCAACAGCACTATATGCGCAACAG 624  
Qy 1741 GCGGGGTACTCCAGAGAGGCTCGAGTCCCATATACAGATGATGTCTCAATGAGGCT 1800  
Db 623 GCGGGGTACTCCAGAGAGGCTCGAGTCCCATATACAGATGATGTCTCAATGAGGCT 564  
Qy 1801 GACTTCTATGAAAACAGATCAAGCCCAAGATGTTCTGTGTGCTTACCCCGAGGTGGC 1860  
Db 563 GACTTCTATGAAAACAGATCAAGCCCAAGATGTTCTGTGTGCTTACCCCGAGGTGGC 504  
Qy 1861 ATTGATGCTGCGCAAGGCGCAAGCGGTGCTTGTGTGTGAGGACAGCATCTCTCGG 1920  
Db 503 ATTGATGCTGCGCAAGGCGCAAGCGGTGCTTGTGTGTGAGGACAGCATCTCTCGG 444  
Qy 1921 ACGCCAGCTTGGGCGGCTGTGTGAGATGTGAGTTGGGGCACTGGCTGTGGCCAG 1980  
Db 443 ACGCCAGCTTGGGCGGCTGTGTGAGATGTGAGTTGGGGCACTGGCTGTGGCCAG 384  
Qy 1981 AAGCCAGCGCTTACACCAAGTCAAGTCACTTCCGAGAGTGAATCTTCAAGCCATTAAG 2040  
Db 383 AAGCCAGCGCTTACACCAAGTCAAGTCACTTCCGAGAGTGAATCTTCAAGCCATTAAG 324  
Qy 2041 ACTCATCTCCGAAGCGGCGCATGTGATCAAGCTGTGACCGGTGCTTCTGTGCGCA 2100  
Db 323 ACTCATCTCCGAAGCGGCGCATGTGATCAAGCTGTGACCGGTGCTTCTGTGCGCA 264  
Qy 2101 GCCTCAGAGGCGGAGGTATCCCGGTGTGGAGTCCACGCTGGGCGAGAGATGGAGCT 2160  
Db 263 GCCTCAGAGGCGGAGGTATCCCGGTGTGGAGTCCACGCTGGGCGAGAGATGGAGCT 204  
Qy 2161 TTTTCTTCTTGGGCGCGGTCACAGATGCTCAAGGACACCTTCCGAGGCTCTCTTTC 2220  
Db 203 TTTTCTTCTTGGGCGCGGTCACAGATGCTCAAGGACACCTTCCGAGGCTCTCTTTC 144  
Qy 2221 CACAGTGGCGGGGCCCATCAGCCCGAGACCAACCACTCAACCTCTTCAACCCCATGT 2280  
Db 143 CACAGTGGCGGGGCCCATCAGCCCGAGACCAACCACTCAACCTCTTCAACCCCATGT 84  
Qy 2281 AAATATTGTTCTGCTGTCTGGAAGTCTGTCTGATAGTCCCGTATGATGAGTCTTTC 2340  
Db 83 AAATATTGTTCTGCTGTCTGGAAGTCTGTCTGATAGTCCCGTATGATGAGTCTTTC 24  
Qy 2341 AAATATTAAAGATGTTTGAAT 2363  
Db 23 AAATATTAAAGATGTTTGAAT 1

RESULT 5  
HSHERSH  
LOCUS 2363 bp mRNA linear PRI 15-JUN-2003  
DEFINITION Human hepatoma mRNA for serine protease hepsin.  
VERSION X07732.1 GI:32063  
KEYWORDS hepsin; membrane protein; serine protease; zymogen.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 2363)  
AUTHORS Leytus,S.P., Loeb,K.R., Hagen,F.S., Kurachi,K. and Davie,E.W.  
TITLE A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells

JOURNAL Biochemistry 27 (3), 1067-1074 (1988)  
MEDLINE 88209431  
PubMed 2835076

COMMENT See X07002 for liver hepsin partial cDNA sequence  
the authors combined the sequence from several overlapping Hep G2  
cDNA clones; additional sequences were found in clones HepG2UW17  
and HepG2UW2.

## FEATURES

Source Location/Qualifiers  
1..2363  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HepG2UW63, HepG2UW61, HepG2UW17, HepG2UW2"  
/cell\_line="Hep G2"  
/tissue\_type="liver"  
/clone\_id="lambda gt11"  
1..191  
/product="HepG2UW7 mRNA"  
83  
/note="5' end of HepG2UW19"  
89  
/note="5' end of HepG2UW61"  
98  
/note="5' end of HepG2UW20"  
116  
/note="5' end of HepG2UW17"  
132  
/note="5' end of HepG2UW63 and HepG2UW2"  
192..771  
/note="Insert in HepG2UW17 (compared to HepG2UW7; put.  
unspliced intron)"  
627..771  
/note="Insert in HepG2UW2 (compared to HepG2UW7; put.  
unspliced intron, part.)"  
772..2363  
/product="HepG2UW7 mRNA"  
826..2079  
/codon\_start=1  
/product="hepsin"  
/protein\_id="CAA30558.1"  
/db\_xref="GI:32064"  
/db\_xref="GOA:P05981"  
/db\_xref="UniProt/Swiss-Prot:P05981"  
/translation="MAQKEGRTVPCSRFVAALVAGTLLTLTAIGASMAIVAVLL  
RSDEPLYPVQVSSADARLWVDTBETWRLCSRSNARVAGSCBEMGRLALVLS  
ELDVRTAGNLSGFPCVDEBRLPHTORLLEIVCDPGRFLAALCODGRALV  
DRIVGSDTSLGRWPMQVSLRYDGAHLGSGLSGDMWLTAAHCFPRNRVLSRWVF  
AGAAOASPHGLDGVAVVYHGVLPDRDPSRNSNDIALVHLSPLPLEYIOPV  
CLPAAGQALVDGKICTYTGKNTQYTGQAGVLDPAKVPITISNDVCSGADPYKQIDP  
KMFAGYPEGGIDACQGDGSGPFCEDSISHTPRWRLGIVSWGTCALAKPKGVYTK  
VSDREWIFQAIKTHSEASGMVTL"  
826..1311  
1312..2076  
/product="unnamed"  
2363

## ORIGIN

Query Match 100.0%; Score 2363; DB 9; Length 2363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCGAGCCGCTTTCCAGGAGACCTAAGAGGCCCCACAGGTGAGGAGCCTGGCTTACG 60  
DB 1 TCGAGCCGCTTTCCAGGAGACCTAAGAGGCCCCACAGGTGAGGAGCCTGGCTTACG 60  
QY 61 AGGCCCCAGCCAGCCTCTGCTCCAGGCGCCCGCTGCTGGGGGCAACATGCTCC 120  
DB 61 AGGCCCCAGCCAGCCTCTGCTCCAGGCGCCCGCTGCTGGGGGCAACATGCTCC 120  
QY 121 TGGCCAGGCTGAGAGCTGACCCGACCCGAGCACTTACGAGGCTCCGCCCACTTGC 180  
DB 121 TGGCCAGGCTGAGAGCTGACCCGACCCGAGCACTTACGAGGCTCCGCCCACTTGC 180

QY 181 TCGACCCAGGGTAAAGGCAAGGGGCCCCAGACTCAAGTTCCAGCCCTGAGGACAGGG 240  
DB 181 TCGACCCAGGGTAAAGGCAAGGGGCCCCAGACTCAAGTTCCAGCCCTGAGGACAGGG 240  
QY 241 TTCCCTCATCCCCCACCACCCAGCCTTAATGCCACCTCTTAATAGAGGGATTCTGGGAGC 300  
DB 241 TTCCCTCATCCCCCACCACCCAGCCTTAATGCCACCTCTTAATAGAGGGATTCTGGGAGC 300  
QY 301 TGAAGAGGGGCACTATGACGTCTCCCAAGACACTTAAAGTGTCTGTCTCTTCTT 360  
DB 301 TGAAGAGGGGCACTATGACGTCTCCCAAGACACTTAAAGTGTCTGTCTCTTCTT 360  
QY 361 CAGACTAGGCGTTGAGACCCAGTCTTCTCTCCAGACCCAGAGATTCCAGGCTCAG 420  
DB 361 CAGACTAGGCGTTGAGACCCAGTCTTCTCTCCAGACCCAGAGATTCCAGGCTCAG 420  
QY 421 GCCCTCCTCCCTCATATCTAGGAGATCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480  
DB 421 GCCCTCCTCCCTCATATCTAGGAGATCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480  
QY 481 TGATTTAGGTCCTCAGCTGTCTCTCCCTCAAAACGGGATCTCAGTCCCTGCTCCAC 540  
DB 481 TGATTTAGGTCCTCAGCTGTCTCTCCCTCAAAACGGGATCTCAGTCCCTGCTCCAC 540  
QY 541 CAGGCTCAGGCGATGGGGGTCCCAATCCCTGCAATCCAGGCGTCCCGGCTGCTGCA 600  
DB 541 CAGGCTCAGGCGATGGGGGTCCCAATCCCTGCAATCCAGGCGTCCCGGCTGCTGCA 600  
QY 601 GACACTAGACCCCATCTTGAACCCAGGCCCAATCTGCTGCTGATCAACAGGCTCTG 660  
DB 601 GACACTAGACCCCATCTTGAACCCAGGCCCAATCTGCTGCTGATCAACAGGCTCTG 660  
QY 661 CCAAGGCGCAGTCCCTTCAAGCTGCTGATGAGCGCTGAGACTGGAGCTGGAGGCGCAGAGACT 720  
DB 661 CCAAGGCGCAGTCCCTTCAAGCTGCTGATGAGCGCTGAGACTGGAGGCGCAGAGACT 720  
QY 721 GGGCTGGGCTGGGCTCCCCCAGGGCTGGCTCCCGGTCATCTCTCAAGGTCACACC 780  
DB 721 GGGCTGGGCTGGGCTCCCCCAGGGCTGGCTCCCGGTCATCTCTCAAGGTCACACC 780  
QY 781 TGGCCCAAGAGGTCAAGCCAGGGAATCAATTAAACAAGAGGAGTGAATGSCCAAGAGAG 840  
DB 781 TGGCCCAAGAGGTCAAGCCAGGGAATCAATTAAACAAGAGGAGTGAATGSCCAAGAGAG 840  
QY 841 GGTGGCCGGAATGTGCAGTCTGTCTCAAGACCCAAGGTGAGCTTCACTGCGGGAGC 900  
DB 841 GGTGGCCGGAATGTGCAGTCTGTCTCAAGACCCAAGGTGAGCTTCACTGCGGGAGC 900  
QY 901 CTGCTACTTCTGACAGGCATCGGGGCGGCAATCTGGGCAATTGTGGCTTTCTCTCAGG 960  
DB 901 CTGCTACTTCTGACAGGCATCGGGGCGGCAATCTGGGCAATTGTGGCTTTCTCTCAGG 960  
QY 961 AGTACAGGAGCGGCTGTATCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
DB 961 AGTACAGGAGCGGCTGTATCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1020  
QY 1021 TTTGACAAAGACGGAAGGAGCGTGGCGGCTGTGTCTTCTGCGCTCAACGCCAGGATA 1080  
DB 1021 TTTGACAAAGACGGAAGGAGCGTGGCGGCTGTGTCTTCTGCGCTCAACGCCAGGATA 1080  
QY 1081 GCGGACTCAGCTGCGAGAGATGGGCTTCTCAGGCACTGACCCACTCCGACTGAGC 1140  
DB 1081 GCGGACTCAGCTGCGAGAGATGGGCTTCTCAGGCACTGACCCACTCCGACTGAGC 1140  
QY 1141 GTGCGAAGCGGCGGCAATGAGCACTGCGGCTTTCTTCTGTGTGAGAGAGGAGGCTG 1200  
DB 1141 GTGCGAAGCGGCGGCAATGAGCACTGCGGCTTTCTTCTGTGTGAGAGAGGAGGCTG 1200  
QY 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCGGTGTGTATGCCCCAGAGGCGGTTTC 1260  
DB 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCGGTGTGTATGCCCCAGAGGCGGTTTC 1260  
QY 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCGGAGAGGCTGCCGTGAGACCGGATGTGTGGGA 1320



Db 1261 TTGGCCGCGCATCTGCGAAGACGTGGCCGCGAAGAGCGCCCGTGGACCGCATGCTGGGA 1320  
Qy 1321 GCGCGGAGACACGACGTTGGGCGCGTGGCCGCTGGCAAGTCAGCTTCCTGCTAATGATGAGACA 1380  
Db 1321 GCGCGGAGACACGACGTTGGGCGCGTGGCCGCTGGCAAGTCAGCTTCCTGCTAATGATGAGACA 1380  
Qy 1381 CACCTCTGTGGGGGATCCCTGCTCTCCGCGGAGCTGGGGCTGACAGACCGCCCACTGCTTC 1440  
Db 1381 CACCTCTGTGGGGGATCCCTGCTCTCCGCGGAGCTGGGGCTGACAGACCGCCCACTGCTTC 1440  
Qy 1441 CCGAGCGGAAACCGGGTCTCTCCGATGAGCGAGTGTTCGCGGTGCGGTGAGCCCAAGCC 1500  
Db 1441 CCGAGCGGAAACCGGGTCTCTCCGATGAGCGAGTGTTCGCGGTGCGGTGAGCCCAAGCC 1500  
Qy 1501 TCTCCCGACCGTCTGCACTGGGGGTGACGCTGTGTCTACACGCGGGGCTATCTTCC 1560  
Db 1501 TCTCCCGACCGTCTGCACTGGGGGTGACGCTGTGTCTACACGCGGGGCTATCTTCC 1560  
Qy 1561 TTTGCGGAGCCCAACGCGGAGAGAACGAAAGATATTTGCGCTGGTCACTCTCCAGT 1620  
Db 1561 TTTGCGGAGCCCAACGCGGAGAGAACGAAAGATATTTGCGCTGGTCACTCTCCAGT 1620  
Qy 1621 CCCCTGCCCCACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGAGGCCCTG 1680  
Db 1621 CCCCTGCCCCACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGAGGCCCTG 1680  
Qy 1681 GTGATGCGAAGATCTGTACCGGTGACGCGGCTGGGGCAACAGCGAGTACTATGCGCAACG 1740  
Db 1681 GTGATGCGAAGATCTGTACCGGTGACGCGGCTGGGGCAACAGCGAGTACTATGCGCAACG 1740  
Qy 1741 GCGCGGGTACTCCAGAGAGGCTCGAGTCCCATTAATCAGCAATGATGTGCAATGAGCCCT 1800  
Db 1741 GCGCGGGTACTCCAGAGAGGCTCGAGTCCCATTAATCAGCAATGATGTGCAATGAGCCCT 1800  
Qy 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCGTACCCGAGGGTGGC 1860  
Db 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCGTACCCGAGGGTGGC 1860  
Qy 1861 ATTATATGCTGCGCAGAGGCGACAGCGGTGTCTCTTGTGTGAGAGACAGATCTTCGG 1920  
Db 1861 ATTATATGCTGCGCAGAGGCGACAGCGGTGTCTCTTGTGTGAGAGACAGATCTTCGG 1920  
Qy 1921 ACGCCAGGTTGGCGGCTGTGTGAGATTGTGAGTTGGGGCACTGGCTGTGGCCCTGAGCCAG 1980  
Db 1921 ACGCCAGGTTGGCGGCTGTGTGAGATTGTGAGTTGGGGCACTGGCTGTGGCCCTGAGCCAG 1980  
Qy 1981 AAGCCAGGCGTCTACCAAAAGTCACTTCGCGGAGTGGATCTTCCAGGCCATTAAG 2040  
Db 1981 AAGCCAGGCGTCTACCAAAAGTCACTTCGCGGAGTGGATCTTCCAGGCCATTAAG 2040  
Qy 2041 ACTACATCCGAAGCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
Db 2041 ACTACATCCGAAGCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100  
Qy 2101 GCTCTCAGAGGCGCAGAGTATCCCGGTGTGATGATGATGATGATGATGATGATGATGATGAT 2160  
Db 2101 GCTCTCAGAGGCGCAGAGTATCCCGGTGTGATGATGATGATGATGATGATGATGATGATGAT 2160  
Qy 2161 TTTTCTTCTTGGGCGCGGTCCACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
Db 2161 TTTTCTTCTTGGGCGCGGTCCACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
Qy 2221 CACGATGCGGCGCCCACTCAGAGCCCGAGACCAACCAACTCAGTCTGATGATGATGATGATGAT 2280  
Db 2221 CACGATGCGGCGCCCACTCAGAGCCCGAGACCAACCAACTCAGTCTGATGATGATGATGATGAT 2280  
Qy 2281 AAATATGTTCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Db 2281 AAATATGTTCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Qy 2341 AAATATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2363  
Db 2341 AAATATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2363

Db 2341 AAATATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2363  
RESULT 6  
CQ718142 1783 bp DNA linear PAT 03-FEB-2004  
LOCUS Sequence 4076 from Patent WO02068579.  
DEFINITION CQ718142  
ACCESSION CQ718142  
VERSION CQ718142.1 GI:42278999  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
JOURNAL Patent: WO 02068579-A 4076 06-SBP-2002;  
FEATURES  
source  
Location/Qualifiers  
1..1783  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 67.4%; Score 1593.6; DB 6; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 2,4e-295;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
Qy 736 CCCCAGGCGCCCTGCTCCCGCTCCATCTCTCAAGGTCACAGCTCCAGCCCGAGAGGTCA 795  
Db 157 CTTGAGGCTTCGCGCCACCTGCTGAGACCCA-GGTTCCACCTCGGACGAGGTCA 215  
Qy 796 GCCAGGAAATATTAAACAGAGGAGTGAATGATGAGCGAGAGAGAGGTGACCGGACTGTG 855  
Db 216 GCCAGGAAATATTAAACAGAGGAGTGAATGATGAGCGAGAGAGAGGTGACCGGACTGTG 275  
Qy 856 CCATGCTGCTCCAGAGCCCAAGGTGAGCTCTCACTGCGGGGACCTTGTCTTACA 915  
Db 276 CCATGCTGCTCCAGAGCCCAAGGTGAGCTCTCACTGCGGGGACCTTGTCTTACA 335  
Qy 916 GCCATCGGGGCGGATCTGAGGCGATGAGGCTGTTCTCTCGAGAGTGAACAGAGCGG 975  
Db 336 GCCATCGGGGCGGATCTGAGGCGATGAGGCTGTTCTCTCGAGAGTGAACAGAGCGG 395  
Qy 976 CTGTACCCAGTGAAGTCACTCTGCGGACGCTCGCTCATGTCTTTGAACAGGAA 1035  
Db 396 CTGTACCCAGTGAAGTCACTCTGCGGACGCTCGCTCATGTCTTTGAACAGGAA 455  
Qy 1036 GGGAGCTGGCGGCTGCTGTCTCTGCGCTCCAGCGGAGGTAAGCTTCAGTTC 1095  
Db 276 GGGAGCTGGCGGCTGCTGTCTCTGCGCTCCAGCGGAGGTAAGCTTCAGTTC 335  
Qy 456 GGGAGCTGGCGGCTGCTGTCTCTGCGCTCCAGCGGAGGTAAGCTTCAGTTC 515  
Db 1096 GAGAGATGAGGCTTCCACAGGAGTGAACCACTCGAGCTGAGCGAGCGAGG 1155  
Qy 1156 GCCAATGAGCAGTCTGAGGCTTCTGTGTGAGAGAGGAGGCTGCCCAACAGAGG 1215  
Db 576 GCCAATGAGCAGTCTGAGGCTTCTGTGTGAGAGAGGAGGCTGCCCAACAGAGG 635  
Qy 1216 CTGTGAGAGTCACTCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1275  
Db 636 CTGTGAGAGTCACTCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 695  
Qy 1276 CAAGACTGTGGCGCGAGGAAGTGGCCGCTGGAACGAGTGTGAGAGGCGGAGACAGAG 1335  
Db 696 CAAGACTGTGGCGCGAGGAAGTGGCCGCTGGAACGAGTGTGAGAGGCGGAGACAGAG 755  
Qy 1336 TTGGGCGGAGTGGCGGAGTCAAGCTTCGCTATGATGAGACACACTCTGTGGGGA 1395  
Db 755 TTGGGCGGAGTGGCGGAGTCAAGCTTCGCTATGATGAGACACACTCTGTGGGGA 1395



Db 756 TTGGGCGGCTGGCGTGGCAATGCTGCTGCTATGATGAGCACTCTGCTGGGGA 815  
 Qy 1396 TCCCTGCTCTCCGGGGAAGCTGGGTGTGACAGCCGCCCACTGCTTCCGGAGCGGAACCG 1455  
 Db 816 TCCCTGCTCTCCGGGGAAGCTGGGTGTGACAGCCGCCCACTGCTTCCGGAGCGGAACCG 875  
 Qy 1456 GTCCGTGCTCCGATGGCGATGCTGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1515  
 Db 876 GTCCGTGCTCCGATGGCGATGCTGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 935  
 Qy 1516 CAGCTGGGGGTGACAGCTGTGTGTCTACACAGGGGGGTCTATCTTCCCTTCCGGAGCGGAAC 1575  
 Db 936 CAGCTGGGGGTGACAGCTGTGTGTCTACACAGGGGGGTCTATCTTCCCTTCCGGAGCGGAAC 995  
 Qy 1576 AGCGAGAGAAAGAGCAAGATATGCTGCTGGGTGCACTCTCCAGTCCCTGCTCCCTGCA 1635  
 Db 996 AGCGAGAGAAAGAGCAAGATATGCTGCTGGGTGCACTCTCCAGTCCCTGCTCCCTGCA 1055  
 Qy 1636 GAATTCATCCAGCTGTGTGTCTCCAGCTGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1695  
 Db 1056 GAATTCATCCAGCTGTGTGTCTCCAGCTGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1115  
 Qy 1696 TGTACCGTGAAGGGCTGGGGGCAACAGAGTATGCTATGAGCAAGGGCGGGGTACTGAG 1755  
 Db 1116 TGTACCGTGAAGGGCTGGGGGCAACAGAGTATGCTATGAGCAAGGGCGGGGTACTGAG 1175  
 Qy 1756 GAGGCTGAGTCCCATATATGAGCAAGTATGCTGCAATGAGCGCTGATCTTATAGAAAC 1815  
 Db 1176 GAGGCTGAGTCCCATATATGAGCAAGTATGCTGCAATGAGCGCTGATCTTATAGAAAC 1235  
 Qy 1816 CAGATCAAGCCCAAGATGCTGT 1875  
 Db 1236 CAGATCAAGCCCAAGATGCTGT 1295  
 Qy 1876 GGCAAGAGCGGT 1935  
 Db 1296 GGCAAGAGCGGT 1355  
 Qy 1936 CTGTGTGAGCATGTTGATGTTGGGGCACTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1995  
 Db 1356 CTGTGTGAGCATGTTGATGTTGGGGCACTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1415  
 Qy 1996 ACCAAAGTCAAGTGAATTCCTGGGAGTGTGATCTTCCAGGCGATTAAGATCTCACTCCGAAGCC 2055  
 Db 1416 ACCAAAGTCAAGTGAATTCCTGGGAGTGTGATCTTCCAGGCGATTAAGATCTCACTCCGAAGCC 1475  
 Qy 2056 AGCGGATGATGAGCCAGCTGTAACCGGTGGCTTCTGCTGGCGGAGCTCCAGGGGCGCGA 2115  
 Db 1476 AGCGGATGATGAGCCAGCTGTAACCGGTGGCTTCTGCTGGCGGAGCTCCAGGGGCGCGA 1535  
 Qy 2116 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2175  
 Db 1536 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1595  
 Qy 2176 CGGTTCACAGGTCCAAGGACACCTCTCCAGGGTCTCTCTTCCAGTGGCGGGGCC 2235  
 Db 1596 CGGTTCACAGGTCCAAGGACACCTCTCCAGGGTCTCTCTTCCAGTGGCGGGGCC 1655  
 Qy 2236 ACTAGAGCGGAGAGCAACCAAGCTCAACCTCTGAGCCGCAATGAAATATGTTCTGCT 2295  
 Db 1656 ACTAGAGCGGAGAGCAACCAAGCTCAACCTCTGAGCCGCAATGAAATATGTTCTGCT 1715  
 Qy 2296 GTCTGAGGATCTCTGTAGT 2355  
 Db 1716 GTCTGAGGATCTCTGTAGT 1775  
 Qy 2356 TTTTGATT 2363  
 Db 1776 TTTTGATT 1783

RESULT 7

AR280046  
 LOCUS AR280046 1783 bp DNA linear PAT 10-APR-2003  
 DEFINITION Sequence 188 from patent US 6518028.  
 ACCESSION AR280046  
 VERSION AR280046.1 GI:29715235  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 1783)  
 O'Brien, T.L.  
 TITLE Methods for the early diagnosis of ovarian and prostate cancer  
 JOURNAL Patent: US 6518028-A 188 11-FEB-2003;  
 FEATURES  
 SOURCE Location/Qualifiers  
 1..1783  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 67.4% Score=1593.67 DB 6; Length 1783;  
 Best Local Similarity 99.1% Pred. No. 2.4e-25;  
 Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 Qy 736 CCCCAGGCGCTGCTCCCGTCCATCTCTCAAGGATCCCAAGGCGGAGGATCA 795  
 Db 157 CTTGAGGCTTCGCCCCCACTGCTGAGACCCA-GGGTCCACCTGGCCAGAGGTCA 215  
 Qy 796 GCCAGGATCATTTAACAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 855  
 Db 216 GCCAGGATCATTTAACAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 275  
 Qy 856 CCATGCTGCTCAGACCCCAAGGTGAGCTCTCACTGCGGGGACCTGCTACTTCTGACA 915  
 Db 276 CCATGCTGCTCAGACCCCAAGGTGAGCTCTCACTGCGGGGACCTGCTACTTCTGACA 335  
 Qy 916 GCCATCGGGGCGGATCCTGGGACCATGTTGGCTGTTCTCTCAGAGATGACAGAGGCG 975  
 Db 336 GCCATCGGGGCGGATCCTGGGACCATGTTGGCTGTTCTCTCAGAGATGACAGAGGCG 395  
 Qy 976 CTGTACCAAGTGAAGTCAAGCTCTGCGAGCGCTGGCTCATGATCTTTGACAGAGCGAA 1035  
 Db 396 CTGTACCAAGTGAAGTCAAGCTCTGCGAGCGCTGGCTCATGATCTTTGACAGAGCGAA 455  
 Qy 1036 GGGAGCTGGGCGCTGCTGTGTCTCTCGGCTCTCAAGCCAGGGATGACCGACTCAGCTGC 1095  
 Db 456 GGGAGCTGGGCGCTGCTGTGTCTCTCGGCTCTCAAGCCAGGGATGACCGACTCAGCTGC 515  
 Qy 1096 GAGGAGATGGGCTTCTCAAGGCACTGACCCACTCCGAGCTGAGAGCTGAGAGCGGCGGGC 1155  
 Db 516 GAGGAGATGGGCTTCTCAAGGCACTGACCCACTCCGAGCTGAGAGCTGAGAGCGGCGGGC 575  
 Qy 1156 GCCAATGAGCAAGTGGGCTTCTTGT 1215  
 Db 576 GCCAATGAGCAAGTGGGCTTCTTGT 635  
 Qy 1216 CTGTGAGGATCATCTCCGT 1275  
 Db 636 CTGTGAGGATCATCTCCGT 695  
 Qy 1276 CAAGACTGTGGCGGAGAGAGCTGCGGTGAGCCGAGATCGTGGAGGCGGGGACACAGC 1335  
 Db 696 CAAGACTGTGGCGGAGAGAGCTGCGGTGAGCCGAGATCGTGGAGGCGGGGACACAGC 755  
 Qy 1336 TTGGGCGGATGGCGGTGAGGATGAGCTTCTGATGATGAGAGCACTCTGTGTGGGGA 1395  
 Db 756 TTGGGCGGATGGCGGTGAGGATGAGCTTCTGATGATGAGAGCACTCTGTGTGGGGA 815  
 Qy 1396 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCGGCCCACTGCTTCCGGAGCGGAACCGG 1455  
 Db 816 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCGGCCCACTGCTTCCGGAGCGGAACCGG 875  
 Qy 1456 GTCTGTCCGATGAGAGTGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1515

Db 876 GTCCTGTCCTCCGAGTGAGTGTGTCGCGGTGCGCGTCCAGGCTCTCCACAGGCTG 935  
Qy 1516 CAGCTGGGGGGTGCAGGCTGTGTCTACTACACGGGGGCTATCTTCCCTTTCGGGACCCCAAC 1575  
Db 936 CAGCTGGGGGGTGCAGGCTGTGTCTACTACACGGGGGCTATCTTCCCTTTCGGGACCCCAAC 995  
Qy 1576 AGCGAGGAGAACACAGATATGACCCTGTGTCACCTCTCAGTCCCTCTCCCTCACA 1635  
Db 996 AGCGAGGAGAACACAGATATGACCCTGTGTCACCTCTCAGTCCCTCTCCCTCACA 1055  
Qy 1636 GAATATCATCCAGCTGTGTGCTCCAGCTGCGCGGCAAGGCTGTGTGATGCGAAGATC 1695  
Db 1056 GAATATCATCCAGCTGTGTGCTCCAGCTGCGCGGCAAGGCTGTGTGATGCGAAGATC 1115  
Qy 1696 TGTACCGGTGAGGGGCTGGGGGCAACAGGATATATGGCCCAAGCGCGGGGTACTCCAG 1755  
Db 1116 TGTACCGGTGAGGGGCTGGGGGCAACAGGATATATGGCCCAAGCGCGGGGTACTCCAG 1175  
Qy 1756 GAGAGCTGAGTCCCATATATCAGCAATGATGTCTGCAATGAGCGCTGACTTCTATGGAAC 1815  
Db 1176 GAGAGCTGAGTCCCATATATCAGCAATGATGTCTGCAATGAGCGCTGACTTCTATGGAAC 1235  
Qy 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTAACCCCGAGGGTGCATTGATGCTGCCAG 1875  
Db 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTAACCCCGAGGGTGCATTGATGCTGCCAG 1295  
Qy 1876 GGCACACGCGGTGCTCTTGTGTGTGAGGACAGCATCTTCGAGCGCCAGCTTGGCGG 1935  
Db 1296 GGCACACGCGGTGCTCTTGTGTGTGAGGACAGCATCTTCGAGCGCCAGCTTGGCGG 1355  
Qy 1936 CTGTGTGGCATTTGTGAGTGGGGGCACTGCTGTGCTGCGCCGCAAGAGCCAGGCTGTAC 1995  
Db 1356 CTGTGTGGCATTTGTGAGTGGGGGCACTGCTGTGCTGCGCCGCAAGAGCCAGGCTGTAC 1415  
Qy 1996 ACCAAGTCAAGTCACTTCGCGGAGTGTGATCTTTCAGGCGCATTAAGTCACTCTCGAAGCC 2055  
Db 1416 ACCAAGTCAAGTCACTTCGCGGAGTGTGATCTTTCAGGCGCATTAAGTCACTCTCGAAGCC 1475  
Qy 2056 AGCGGATGTGTGACCCAGCTCTGACCGGTGCTTCTGCTGCGGAGCGCTCCAGGCGCGGA 2115  
Db 1476 AGCGGATGTGTGACCCAGCTCTGACCGGTGCTTCTGCTGCGGAGCGCTCCAGGCGCGGA 1535  
Qy 2116 GGTATCCCGGTGTGTGGGATCCAGCTGTGGGCGGAGATGGGACGTTTTCTTCTTGGGCGC 2175  
Db 1536 GGTATCCCGGTGTGTGGGATCCAGCTGTGGGCGGAGATGGGACGTTTTCTTCTTGGGCGC 1595  
Qy 2176 CGGTCCAGGTCCAGAGACACCTCTCCCTCAGGGGTCTCTTTCACAGTGGCGGGCGCC 2235  
Db 1596 CGGTCCAGGTCCAGAGACACCTCTCCCTCAGGGGTCTCTTTCACAGTGGCGGGCGCC 1655  
Qy 2236 ACTACGCCCCGAGACCAACCCCACTCAACCTCTCGACCCCAATGTAATATTTGTTGCT 2295  
Db 1656 ACTACGCCCCGAGACCAACCCCACTCAACCTCTCGACCCCAATGTAATATTTGTTGCT 1715  
Qy 2296 GTCTGGAGCTCTGTATAGTGTCCCTGATGATGGGATGCTTTAATAATAAAGATGG 2355  
Db 1716 GTCTGGAGCTCTGTATAGTGTCCCTGATGATGGGATGCTTTAATAATAAAGATGG 1775  
Qy 2356 TTTTGATT 2363  
Db 1776 TTTTGATT 1783

RESULT 8  
AX207963 1783 bp DNA linear PAT 31-AUG-2001  
LOCUS AX207963  
DEFINITION Sequence 67 from Patent WO0157194.  
ACCESSION AX207963  
VERSION AX207963.1 GI:15422559  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Butheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Madison, E.L., Ong, F.O. and Yeh, J.C.  
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the encoded proteins and methods based thereon  
JOURNAL Patent: WO 0157194-A 67 09-AUG-2001;  
CORVAS INTERNATIONAL, INC. (US)  
FEATURES  
source Location/Qualifiers  
1..1783  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
246..1499  
/note="unassigned protein product; Nucleotide sequence encoding human hepsin"  
/codon\_start=1  
/protein\_id="CA60387.1"  
/db\_xref="GI:15422560"  
/translation="MAQKEGRVPPCSRPKVALTAGTLLTLTAIGASWAIWAVLL  
RSDPELVYPOVSADARLWVFDKTEGTWRILCSRSNARVAGLSEEMFGLALTHS  
ELDVRTAGANGTSGFPCVDRGLPHORLLEVLVSDCPGRFLAICODGGRKL  
DRIVGRDTSLEKRPWQVSLRTDGAHLGSSLLSGDWVLTAAHCFEPRNRYLRWVF  
AGVVAQSPHGLQIVQAVVYHGGYLPFRDPSNENSDIALVHLSPLFTEYIDP  
CLPAAGQALVDKICTVTGWNTOYQOAGVQEARVPIISDVCNADGFYNOJIKP  
KMFCAQPEEGIDACODSGSPVCEDSISRTPRMRICGIVSWGTCAALAKKGVYTK  
VSDFRMIFQAIKTHSBASGMVQL"

ORIGIN  
Query Match 67.4%; Score 1593.6; DB 6; Length 1783;  
Best Local Similarity 99.1%; Pred. No. 2,4e-295;  
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
Qy 736 CCCCAGGCGCTGCTCCCGCTCCATCTCTCAAGGTCCACCGCTGGGCGCCAGAGGTCA 795  
Db 157 CTTGAGGCTCCCGCCCACTCTGTGAGCCCA-GGTCCACCTGTGGCCAGAGGTCA 215  
Qy 796 GCAGAGAAATCATTAACAGAGGAGTGAATGACGCGAGAGGAGGGTGGCCGACCTGTG 855  
Db 216 GCAGAGAAATCATTAACAGAGGAGTGAATGACGCGAGAGGAGGGTGGCCGACCTGTG 275  
Qy 856 CCAATGCTCTCAGACCCAGAGTGGGACCTCTCACTGCGGGGACCTGCTACTTCTACA 915  
Db 276 CCAATGCTCTCAGACCCAGAGTGGGACCTCTCACTGCGGGGACCTGCTACTTCTACA 335  
Qy 916 GCCATGGGGGCGCATCTGTGGGCAATGAGGCTGTTCTCTCGAGAGTCAAGAGCGG 975  
Db 336 GCCATGGGGGCGCATCTGTGGGCAATGAGGCTGTTCTCTCGAGAGTCAAGAGCGG 395  
Qy 976 CCAATGAGGAGTCACTCTGCGGACGCTCGGCTCAATGATCTTTGAACAAGGAA 1035  
Db 396 CCAATGAGGAGTCACTCTGCGGACGCTCGGCTCAATGATCTTTGAACAAGGAA 455  
Qy 1036 GGAAGCTGGGCTGTGTCTCTGCGCTCCAGCGCCAGGGTACCGGACTCACTGTC 1095  
Db 456 GGAAGCTGGGCTGTGTCTCTGCGCTCCAGCGCCAGGGTACCGGACTCACTGTC 515  
Qy 1096 GAGAGATGGGCTTCCAGAGGAGTGAACCATCCGAGGTGAGACGCGGAGCGGCGG 1155  
Db 516 GAGAGATGGGCTTCCAGAGGAGTGAACCATCCGAGGTGAGACGCGGAGCGGCGG 575  
Qy 1156 GCCAATGGACATCTCGGCTTCTCTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 1215  
Db 576 GCCAATGGACATCTCGGCTTCTCTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 635  
Qy 1216 CTGCTGGAGGTCACTTCCTGTGTGATGCTCCAGAGGCGGTTTCTTGGCGCATCTGC 1275  
Db 636 CTGCTGGAGGTCACTTCCTGTGTGATGCTCCAGAGGCGGTTTCTTGGCGCATCTGC 695  
Qy 1276 CAAGACTGTGGCGCGAGAGAGTGGCCGTGAGACCGATGCTGGAGAGGCGGAGACCCAGC 1335  
Db 696 CAAGACTGTGGCGCGAGAGAGTGGCCGTGAGACCGATGCTGGAGAGGCGGAGACCCAGC 755  
Qy 1336 TTGGGCGGAGCGGAGGCAAGTCAAGCTTCCGTATGATGAGACACACTCTGTGGGGA 1395

|    |      |  |      |
|----|------|--|------|
| Db | 756  | TTGGGCGGATGGCCGTGGCAAGTCAGCCTTCGCTATGATGGAGCAACACTCTGTGGGGGA   | 815  |
| Qy | 1396 | TCCCTGCTCTCCGGGGACTGGGTGCTGACAGCCGCCCACTGCTTCCGGAGCGAAACGG     | 1455 |
| Db | 816  | TCCCTGCTCTCCGGGGACTGGGTGCTGACAGCCGCCCACTGCTTCCGGAGCGAAACGG     | 875  |
| Qy | 1456 | GTCCTGTCCGATGGCGATGTTTGGCCGGTGGCCGGTGGCCAGAGCCTCTCCCAAGGTCTG   | 1515 |
| Db | 876  | GTCCTGTCCGATGGCGATGTTTGGCCGGTGGCCGGTGGCCAGAGCCTCTCTCCCAAGGTCTG | 935  |
| Qy | 1516 | CAGCTGGGGGTGCAAGCTGTGGTCTACACAGGGGGCTATCTCTCCCTTCGGGACCCCAAC   | 1575 |
| Db | 936  | CAGCTGGGGGTGCAAGCTGTGGTCTACACAGGGGGCTATCTCTCCCTTCGGGACCCCAAC   | 995  |
| Qy | 1576 | AGCGAGAGAAACAGACGATATTGCTCCCTGCTCAACTCTTCCAGTCCCTGCTCCACA      | 1635 |
| Db | 996  | AGCGAGAGAAACAGACGATATTGCTCCCTGCTCAACTCTTCCAGTCCCTGCTCCACA      | 1055 |
| Qy | 1636 | GAATATCATCCAGCCGTGTGGCTCCCACTGCGCGGCGAGGGCCGTGGTGGATGGCAAGATC  | 1695 |
| Db | 1056 | GAATATCATCCAGCCGTGTGGCTCCCACTGCGCGGCGAGGGCCGTGGTGGATGGCAAGATC  | 1115 |
| Qy | 1696 | TGTAACCTGACGGGCTGGGGCAACAGAGTACTATGGCCAAAGAGCCGGGGTACTCGAG     | 1755 |
| Db | 1116 | TGTAACCTGACGGGCTGGGGCAACAGAGTACTATGGCCAAAGAGCCGGGGTACTCGAG     | 1175 |
| Qy | 1756 | GAGGCTCGAGTCCCATATCAGCAATGATGTCTGCAATGGCGCTGACTTCTATGAAAC      | 1815 |
| Db | 1176 | GAGGCTCGAGTCCCATATCAGCAATGATGTCTGCAATGGCGCTGACTTCTATGAAAC      | 1235 |
| Qy | 1816 | CAGATCAAGCCCAAGATTTCTGTGCTGGCTTACCCCGAGGGTGGATTTGATGCCCTGCCAG  | 1875 |
| Db | 1236 | CAGATCAAGCCCAAGATTTCTGTGCTGGCTTACCCCGAGGGTGGATTTGATGCCCTGCCAG  | 1295 |
| Qy | 1876 | GCGCAGAGCGGTGGTCCCTTGTGTGTAGAGACAGCATCTCTCGGACGCCAGTGTGGCGG    | 1935 |
| Db | 1296 | GCGCAGAGCGGTGGTCCCTTGTGTGTAGAGACAGCATCTCTCGGACGCCAGTGTGGCGG    | 1355 |
| Qy | 1936 | CTGTGTGGCATTTGTGATTTGGGGGCACTGGCTGTGCTTGGCCCGAGAGCCAGGCGTTAC   | 1995 |
| Db | 1356 | CTGTGTGGCATTTGTGATTTGGGGGCACTGGCTGTGCTTGGCCCGAGAGCCAGGCGTTAC   | 1415 |
| Qy | 1996 | ACCAAGTCAGTGAATTCCGGGAGTGGATCTTTCAGGGCCATTAAGAACCTCACTCCGAGACC | 2055 |
| Db | 1416 | ACCAAGTCAGTGAATTCCGGGAGTGGATCTTTCAGGGCCATTAAGAACCTCACTCCGAGACC | 1475 |
| Qy | 2056 | AGCGGCAATGTGACCCAGACTCTGACCCGGTGGCTTCTCGCTCGAGCCAGCTCCAGGGCCGA | 2115 |
| Db | 1476 | AGCGGCAATGTGACCCAGACTCTGACCCGGTGGCTTCTCGCTCGAGCCAGCTCCAGGGCCGA | 1535 |
| Qy | 2116 | GGTGATCCCGGTGGTGGATCCAGCTGGGCCAGAGATGGAGCGTTTCTTCTTGGGCC       | 2175 |
| Db | 1536 | GGTGATCCCGGTGGTGGATCCAGCTGGGCCAGAGATGGAGCGTTTCTTCTTGGGCC       | 1595 |
| Qy | 2176 | CGGTGCCAAGGTCCCAAGGACACCCCTCCCTCCAGGGTCCCTCTCTTCCAGAGTGGGGGCC  | 2235 |
| Db | 1596 | CGGTGCCAAGGTCCCAAGGACACCCCTCCCTCCAGGGTCCCTCTCTTCCAGAGTGGGGGCC  | 1655 |
| Qy | 2236 | ACTCAGCCCGGAGACCAACCAACTCAACCTCTCTGACCCCGCATGTAAATATTGTCTGCT   | 2295 |
| Db | 1656 | ACTCAGCCCGGAGACCAACCAACTCAACCTCTCTGACCCCGCATGTAAATATTGTCTGCT   | 1715 |
| Qy | 2296 | GTCGTGGACTCTGTCTTAGTGGCCCTGATGTATGGAGTGTCTTTAAATATAAAGATGG     | 2355 |
| Db | 1716 | GTCGTGGACTCTGTCTTAGTGGCCCTGATGTATGGAGTGTCTTTAAATATAAAGATGG     | 1775 |
| Qy | 2356 | TTTGGATT 2363  |      |
| Db | 1776 | TTTGGATT 1783  |      |

| LOCUS                      | 1783 bp  | mRNA                         | linear | PRI 11-JUN-1993 |
|----------------------------|--|------------------------------|--------|-----------------|
| DEFINITION                 | Human hepsin mRNA, complete cds.   |                              |        |                 |
| ACCESSION                  | M18930   |                              |        |                 |
| VERSION                    | M18930.1   |                              |        |                 |
| KEYWORDS                   | GI:184371  |                              |        |                 |
| SOURCE                     | hepsin; protease; serine protease.   |                              |        |                 |
| ORGANISM                   | Homo sapiens (human)   |                              |        |                 |
| REFERENCE                  | Homo sapiens   |                              |        |                 |
| AUTHORS                    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.    |                              |        |                 |
| TITLE                      | Leytus, S.P., Loeb, K.R., Hagen, F.S., Kurchi, K. and Davie, E.W.  |                              |        |                 |
| JOURNAL                    | A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells |                              |        |                 |
| MEDLINE                    | Biochemistry 27 (3), 1067-1074 (1988)  |                              |        |                 |
| PUBMED                     | 88209431   |                              |        |                 |
| COMMENT                    | 2835076  |                              |        |                 |
| FEATURES                   | Original source text: Human liver cell, cDNA to mRNA, clone HepG2UW7.  |                              |        |                 |
| SOURCE                     | Location/Qualifiers  |                              |        |                 |
|                            | 1..1783  |                              |        |                 |
|                            | /organism="Homo sapiens"   |                              |        |                 |
|                            | /mol_type="mRNA"   |                              |        |                 |
|                            | /db_xref="taxon:9606"  |                              |        |                 |
|                            | 1..1783  |                              |        |                 |
|                            | /product="hepsin mRNA"   |                              |        |                 |
|                            | 246..1499  |                              |        |                 |
|                            | /note="hepsin (serine protease) precursor"   |                              |        |                 |
|                            | /codon_start=1   |                              |        |                 |
|                            | /protein_id="AA36013.1"  |                              |        |                 |
|                            | /db_xref="GI:306886"   |                              |        |                 |
|                            | /translation="MAQKEGRTVPCSPRYAAALTAGTLLTLTAIGASWAIVAVLL  |                              |        |                 |
|                            | RSDEPLTPVQVSSADARLWVDFDTBPTWLLGSSRSNARVAGISCMEGFLALHTHS  |                              |        |                 |
|                            | ELDVRTAGANGTSGPCVDEBGLPHTGTLLEIVICDPCRPRFLAALICDGRRLKPL  |                              |        |                 |
|                            | DRVAGRDTSLGRPMQVSLRYDGAALCGSLSGDMVLTAAHCFERNRLSRVRY  |                              |        |                 |
|                            | AGVAAQSLSPGILQGNQAVVYHGALPPRPNSENNDIALVHLSPLTLEYIOPV   |                              |        |                 |
|                            | CLPAAGALVDGKICTYTGMENTPTYYGQAGVLOBAVPIISNVCSADPYGNQIKR   |                              |        |                 |
|                            | KMFQAVYBEGLDAGDSDGPFVCEISIRTPMRRLGIVSGTGALQKPKPVYTRK   |                              |        |                 |
|                            | VSDPFWIFQAIKTHSASGMYTLQ"   |                              |        |                 |
|                            | 246..731   |                              |        |                 |
|                            | /note="hepsin signal peptide"  |                              |        |                 |
|                            | 732..1496  |                              |        |                 |
|                            | /product="hepsin"  |                              |        |                 |
|                            | 297..377   |                              |        |                 |
|                            | /note="transmembrane domain"   |                              |        |                 |
|                            | ORIGIN   | 35 bp upstream of Apal site. |        |                 |
| Query Match                | 67.4%; Score 1593.6; DB 9; Length 1783;  |                              |        |                 |
| Best Local Similarity      | 99.1%; Pred. No. 2.4e-29;  |                              |        |                 |
| Matches 1613; Conservative | 0; Mismatches 14; Indels 1; Gaps 1;  |                              |        |                 |
| QY                         | 736 CCCCCAGGCCCTGGCTCCCGCTCATCTCCCTCAGACAGGTCCCAACCTGGACCCAGAGGTCA   | 795                          |        |                 |
| DB                         | 157 CCTGAGAGCTCCGCCCACTGCTGACACCCA-GGTTCCACCTTGCCACAGAGGTCA  | 215                          |        |                 |
| QY                         | 796 GCCAGGGAATCATTTAACAGAGGCGAGTGAATGCGGAGAAAGAGGTTGCCGAGCTGTG   | 855                          |        |                 |
| DB                         | 216 GCCAGGGAATCATTTAACAGAGGCGAGTGAATGCGGAGAAAGAGGTTGCCGAGCTGTG   | 275                          |        |                 |
| QY                         | 856 CCATGCTGCTCCAAACCCAGGTGGCAGCTTCACTGCGGGGACCTTGCTCTTCTGACA  | 915                          |        |                 |
| DB                         | 276 CCAATGCTGCTCCAAACCCAGGTGGCAGCTTCACTGCGGGGACCTTGCTCTTCTGACA   | 335                          |        |                 |
| QY                         | 916 GCATCGGGGCGGATCTTGAGGCATTTGGCTTTCTCCACAGAGTGAACAGAGCGG   | 975                          |        |                 |
| DB                         | 336 GCATCGGGGCGGATCTTGAGGCATTTGGCTTTCTCCACAGAGTGAACAGAGCGG   | 395                          |        |                 |
| QY                         | 976 CTGTACCAAGTGACAGTCAAGCTCTGCGGACGCTCGGCTCATGTGCTTTGACAAAGCGAA   | 1035                         |        |                 |
| DB                         | 396 CTGTACCAAGTGACAGTCAAGCTCTGCGGACGCTCGGCTCATGTGCTTTGACAAAGCGAA   | 455                          |        |                 |
| QY                         | 1036 GGAACGTGGCGGCTGCTGTGCTCTCGCGCTTCAACGCCAGAGTGAACCGAATCAAGCTGC  | 1095                         |        |                 |

|    |      |   |      |
|----|------|---|------|
| Db | 456  | GGGACGTGCGCGCTCTGTGCTCTTCGCGCTTCAAAGCCAGGGTACGGGACTCAGCTGC        | 515  |
| Oy | 1096 | GAGGAGATGAGCTCTCTCAGGGCACTGACCACTCCAGCTTGAGCTGACGACCGGCGGC        | 1155 |
| Db | 516  | GAGGAGATGAGGCTTCTCTCAGGGGCACTGAACCACTCCAGAGCTTGAGCTGTGAAACCGCGGGC | 575  |
| Oy | 1156 | GCCATGAGCACTCGGAGCTTCTTCTGTGTGAGACGAGGGGAGGGCTGCCCAACCCAGAG       | 1215 |
| Db | 576  | GCATATGGCAAGTCGGAGCTTCTTCTGTGTGAGACGAGGGGAGGGCTGCCCAACCCAGAG      | 635  |
| Oy | 1216 | CTGCTGAGAGGTATCATCTCCGTGTGATTTGCCCCAAGAGGCGGTTTCTTGAGCGCCATCTGC   | 1275 |
| Db | 636  | CTGCTGAGAGGTATCTCCGTGTGTATTTGCCCCAGAGGCGGTTTCTTGAGCGCCATCTGC      | 695  |
| Oy | 1276 | CAGAAGCTGTGGCCGAGGAAAGCTGCGCGTGAACCGCATGTGTGGAGGCGGGACACAGC       | 1335 |
| Db | 696  | CAAGAAGCTGTGGCCGAGGAAAGCTGCGCGTGAACCGCATGTGTGGAGGCGGGACACAGC      | 755  |
| Oy | 1336 | TTGGGACCGGTGCGCCGTGCGAAGTCAAGCTTGCGCTATGATGAGACACACTCTGTGTGGGGA   | 1395 |
| Db | 756  | TTGGGACCGGTGCGCCGTGCGAAGTCAAGCTTGCGCTATGATGAGACACACTCTGTGTGGGGA   | 815  |
| Oy | 1396 | TCCCGCTCTCCGCGGGGACGAGGTGCTGACAGCGCGCCCATCTGCTCCGAGGAGAAACGG      | 1455 |
| Db | 816  | TCCCGCTCTCCGCGGGGACGAGGTGCTGACAGCGCGCCCATCTGCTCCGAGGAGAAACGG      | 875  |
| Oy | 1456 | GTCTGTCTCCGATGAGCGAGTGTGTTGGCCGGGTGCGGTGAGCCAGGCTCTCCCAAGGTCTG    | 1515 |
| Db | 876  | GTCTGTCTCCGATGAGCGAGTGTGTTGGCCGGGTGCGGTGAGCCAGGCTCTCCCAAGGTCTG    | 935  |
| Oy | 1516 | CAGCTGGGGGGGTGACAGGCTGTGTCTTACCAAGGGGGCTATCTTCCCTTTCCGGACCCCAAC   | 1575 |
| Db | 936  | CAGCTGGGGGGGTGACAGGCTGTGTCTTACCAAGGGGGCTATCTTCCCTTTCCGGACCCCAAC   | 995  |
| Oy | 1576 | AGCGAGGAGAACAGAACGATATTTGGCCGTGTCACCTCCAGTCCGCTGCCCCCTCA          | 1635 |
| Db | 996  | AGCGAGGAGAACAGAACGATATTTGGCCGTGTCACCTCCAGTCCGCTGCCCCCTCA          | 1055 |
| Oy | 1636 | GAATATCATTCAGCTGTGTGCTCTCCAGCTGCGCGACAGGCTTGATGATGACAAAGTC        | 1695 |
| Db | 1056 | GAATATCATTCAGCTGTGTGCTCTCCAGCTGCGCGACAGGCTTGATGATGACAAAGTC        | 1115 |
| Oy | 1696 | TGTACCTGACCGGGCTGGGGCAACAGCAGTACTATGACCAACAGGCGGGGTACTCAG         | 1755 |
| Db | 1116 | TGTACCTGACCGGGCTGGGGCAACAGCAGTACTATGACCAACAGGCGGGGTACTCAG         | 1175 |
| Oy | 1756 | GAGGCTGAGTCCCATATATACGAATATGTCTGCATATGCGCTGACTTCTATAGGAAC         | 1815 |
| Db | 1176 | GAGGCTGAGTCCCATATATACGAATATGTCTGCATATGCGCTGACTTCTATAGGAAC         | 1235 |
| Oy | 1816 | CAGATCAAGCCCAAGATGTTCTGTGCTGAGTACCCGAGGGGTGATGATGCTGACAG          | 1875 |
| Db | 1236 | CAGATCAAGCCCAAGATGTTCTGTGCTGAGTACCCGAGGGGTGATGATGCTGACAG          | 1295 |
| Oy | 1876 | GAGCAGACGGGTGTCTCCTTTGTGTGTGAGGACAGCAATCTCTCGACGCGCAGTTGGCGG      | 1935 |
| Db | 1296 | GAGCAGACGGGTGTCTCCTTTGTGTGTGAGGACAGCAATCTCTCGACGCGCAGTTGGCGG      | 1355 |
| Oy | 1936 | CTGTGTGCAATTTGTGATTTGGGGGCACTGAGCTGTGCTCGGGCCCAAGACCGCTCTAC       | 1995 |
| Db | 1356 | CTGTGTGCAATTTGTGATTTGGGGGCACTGAGCTGTGCTCGGGCCCAAGACCGCTCTAC       | 1415 |
| Oy | 1996 | ACCAATGTCAGTCACTTCCGGGAGTGAATCTTCCAGGCAATPAAGACCTCACCTCCGAGCC     | 2055 |
| Db | 1416 | ACCAATGTCAGTCACTTCCGGGAGTGAATCTTCCAGGCAATPAAGACCTCACCTCCGAGCC     | 1475 |
| Oy | 2056 | AGCGGCAATGTGACCCAGCTGTGAACCGGTGCTTCTGCTGCGACGCTCCAGGGCCCGA        | 2115 |
| Db | 1476 | AGCGGCAATGTGACCCAGCTGTGAACCGGTGCTTCTGCTGCGACGCTCCAGGGCCCGA        | 1535 |
| Oy | 2116 | GCTGATCCCGGTGTGTGGATTCACGCTGTGGCCGAGAGTGGACGTTTTTCTTCTTGGGCC      | 2175 |
| Db | 1536 | GCTGATCCCGGTGTGTGGATTCACGCTGTGGCCGAGAGTGGACGTTTTTCTTCTTGGGCC      | 1595 |

| QY         | 2176  | CGGTCCACAGGTCCAAAGACACCTTCCCTCCAGGGTCTCTCTTTCCACAGTGGCGGGGCC    | 2235                        |
|------------|---|---|-----------------------------|
| Db         | 1596  | CGGTCCACAGGTCCAAAGACACCTTCCCTCCAGGGTCTCTCTTTCCACAGTGGCGGGGCC    | 1655                        |
| QY         | 2236  | ACTAGAGCCCGAGACCAACCCAACTCAACCTCTCTGACCCCGCATGTAAATATTTGTTCTGCT | 2295                        |
| Db         | 1656  | ACTAGAGCCCGAGACCAACCCAACTCAACCTCTCTGACCCCGCATGTAAATATTTGTTCTGCT | 1715                        |
| QY         | 2296  | GTTTGGGACTCTGTCTAGGTGCCCCCTGATGATGGATGCTCTTTAAATTAATAAAGATGG    | 2355                        |
| Db         | 1716  | GTTTGGGACTCTGTCTAGGTGCCCCCTGATGATGGATGCTCTTTAAATTAATAAAGATGG    | 1755                        |
| QY         | 2356  | TTTTGATT 2363   |                             |
| Db         | 1776  | TTTTGATT 1783   |                             |
| RESULT 10  |   |   |                             |
| LOCUS      | BC025716  |   |                             |
| DEFINITION | BC025716  | 1761 bp   | mRNA linear PRI 29-JUN-2004 |
| ACCESSION  | BC025716  |   |                             |
| VERSION    | BC025716.1  |   | GI:19343934                 |
| KEYWORDS   | MGC.  |   |                             |
| SOURCE     | Homio sapiens (human)   |   |                             |
| ORGANISM   | Homio sapiens   |   |                             |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |   |                             |
| AUTHORS    | Strausberg, R. L., Felingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Sherman, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Ditchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Schaefer, T. E., Brownstein, M. J., Udell, T. B., Toshitsuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullany, S. J., Bosak, S. A., McEwan, P. J., McEran, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hilyk, S. W., Vallalun, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, J. A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bonfield, W. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butlerfield, Y. S., Krzywinski, M. I., Skalska, U. J., Smalins, D. E., Scherch, A., Schein, J. E., Jones, S. J., and Marz, M. A. |   |                             |
| TITLE      | Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  |   |                             |
| JOURNAL    | Proc. Natl. Acad. Sci. U.S.A.   | 99 (26),  | 16899-16903 (2002)          |
| PUBMED     | 12477932  |   |                             |
| REFERENCE  | 2 (bases 1 to 1761)   |   |                             |
| AUTHORS    | Strausberg, R.  |   |                             |
| TITLE      | Direct Submission   |   |                             |
| JOURNAL    | Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  |   |                             |
| REMARK     | NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>  |   |                             |
| COMMENT    | Contact: MGC help desk<br>Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a><br>Tissue Procurement: Life Technologies, Inc.<br>cDNA Library Preparation: Life Technologies, Inc.<br>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)<br>DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),<br>Gaithersburg, Maryland;<br>Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a><br>Contact: <a href="mailto:nisc_mgc@ncl.nih.gov">nisc_mgc@ncl.nih.gov</a><br>Ahter, N., Ayle, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakeley, R. W., Bonfield, G. G., Breen, K., Binkley, C., Brooks, S.,  |   |                             |

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stankitop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 49 Row: 0 Column: 11  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504480.

## FEATURES

## Source

Location/Qualifiers

1..1761

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:34588 IMAGE:5228525"

/tissue\_type="Pancreas, Spleen, adult pooled"

/clone\_id="NIH\_MGC\_120"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..1761

/gene="HPN"

/note="Synonym: TPNRSS1"

/db\_xref="locusID:3249"

/db\_xref="MIM:142440"

198..1451

/gene="HPN"

/codon\_start=1

/product="Hepsin (transmembrane protease, serine 1)"

/protein\_id="AAH25716.1"

/db\_xref="GI:19343935"

/db\_xref="locusID:3249"

/db\_xref="MIM:142440"

/translation="MAQEGRTVPCCSRPVVAATAGTLLILTAIGASWAVVLLRSDEPLYPVOVSADARLNFVDTBETWRLCSSRSNARVAGISCEMGLRLTSS

ELDVRTAGANGTSGLFCVDEGRLEPHOTRLLEIVSCCPREFLAALCDDGRKLV

DRIVGRDTSIGRMVQVSLRYDGAHLGGSLVLSGDWVLTAAHFPERRNLTSMRYF

AGVAOASPHGLQGVQVAVVYHGYLFPDRDSEBNSNDIALVHLSPLPLEYIOV

CLPAGQALVNDGKICVTLMGNTQYGGQGVQLQEARVITISNDVCNADYGNQIDP

KMPCAGYBEGSIDACGDSGSPRYCEPDSISRTPRMRLCGIVSNQTCALAKQPEVYTK

VSDREMLFQAIKTHSEASGMVTL"

## ORIGIN

Query Match 67.4%; Score 1592; DB 9; Length 1761;

Best Local Similarity 99.0%; Pred. No. 4,9e-295;

Matches 1612; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

736 CCCCAGGCTCTGCTCCCGCTCCATCTCTCAAGGTCCACCTGGCCAGAGGTCA 795  
109 CCTGAGGCTCGCGCCCACTGCTGACCCCA-GGGTCCCACTGGCCAGAGGTCA 167  
796 GCCAGGGAATCATTAACAAGAGAGTGAATGCGCGAAGAGAGGTGGCGGACTGTG 855  
168 GCCAGGGAATCATTAACAAGAGAGTGAATGCGCGAAGAGAGGTGGCGGACTGTG 222  
856 CCATGCTCTCAGACCCCAAGGTGAGCTCTCACTGGGGGACCTGCTCTCTGACA 915  
228 CCACTGCTCTCAGACCCCAAGGTGAGCTCTCACTGGGGGACCTGCTCTCTGACA 287  
916 GCCATCGGGGGGAGCATCTGCGGCAATTTGGCTGTTCTCTCAGAGTGAACAGAGCGG 975  
288 GCCATCGGGGGGAGCATCTGCGGCAATTTGGCTGTTCTCTCAGAGTGAACAGAGCGG 347  
976 CTGTAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1035  
348 CTGTAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 407  
1036 GGGAGTGGCGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1095  
408 GGGAGTGGCGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 467

1096 GAGAGATATGGCTTCTCTCAAGGCACTGACCCACTCCAGGCTGAGCTGCGAAGCGGGGC 1155  
468 GAGAGATATGGCTTCTCTCAAGGCACTGACCCACTCCAGGCTGAGCTGCGAAGCGGGGC 527  
1156 GCCAATATGAGCTGAGGCTTCTCTGATGAGCAAGAGGAGGCTGCCCAACACGAGAGG 1215  
528 GCCAATATGAGCTGAGGCTTCTCTGATGAGCAAGAGGAGGCTGCCCAACACGAGAGG 587  
1216 CTGCTGAGAGTGAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1275  
588 CTGCTGAGAGTGAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 647  
1276 CAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1335  
648 CAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 707  
1336 TTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1395  
708 TTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 767  
1396 TCCCTGCTCTCTCGGAGAGCTGAGTGAACAGCGGCCCACTGCTTCCGAGCGGAACCGG 1455  
768 TCCCTGCTCTCTCGGAGAGCTGAGTGAACAGCGGCCCACTGCTTCCGAGCGGAACCGG 827  
1456 GTCTGCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1515  
828 GTCTGCTCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 887  
1516 CAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1575  
888 CAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 947  
1576 AGCAG 1635  
948 AGCAG 1007  
1636 GAATACATCCAGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1695  
1008 GAATACATCCAGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1067  
1696 TGTACGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1755  
1068 TGTACGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1127  
1756 GAGGCTCAGATCCCATTAATCAAGATGATGATGATGATGATGATGATGATGATGATG 1815  
1128 GAGGCTCAGATCCCATTAATCAAGATGATGATGATGATGATGATGATGATGATGATG 1187  
1816 CAGATCAAGCCCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1875  
1188 CAGATCAAGCCCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1247  
1876 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1935  
1248 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1307  
1936 CTGCTGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1995  
1308 CTGCTGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1367  
1996 ACCAAGTCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2055  
1368 ACCAAGTCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1427  
2056 AGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2115  
1428 AGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1487  
2116 GGTGATCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2175  
1488 GGTGATCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1547  
2176 CGGTCCACAGGTCCACAGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2235



|   |  |   |        |    |      |   |      |
|---|--|---|--------|----|------|---|------|
| Db  | 1548                                   | CGGTCACAGAGTCCAGAGACACCTCCCTCCAGGGTCCCTCTTCCACAGTGGGGGCC    | 1607   | Db | 421  | GGAGAGCTGCCCAACCCAGAGCTGTGAGGTATCTCCGTGTGTATGTCGCCAGA       | 480  |
| Qy  | 2236                                   | ACTCAGCCCCGAGACCAACCACTCACCTCTCTGAGCCCACTGAATATTTGCT        | 2235   | Qy | 1252 | GGCCGTTCTTTGGCCCGCATCTGCCA-----                             | 1277 |
| Db  | 1608                                   | ACTCAGCCCCGAGACCAACCACTCACCTCTCTGAGCCCACTGAATATTTGCT        | 1667   | Db | 481  | GGCCGTTCTTTGGCCCGCATCTGCCAAGTGAATCTTAAATCAGAACCTCTCTT       | 540  |
| Qy  | 2296                                   | GTCTGAGAGCTCCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG        | 2355   | Qy | 1278 | -----   | 1277 |
| Db  | 1668                                   | GTCTGAGAGCTCCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG        | 1727   | Db | 541  | AGGCGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG       | 600  |
| Qy  | 2356                                   | TTTGTATT 2363   |        | Qy | 1278 | -----   | 1277 |
| Db  | 1728                                   | TTTGTATT 1735   |        | Db | 601  | CCCCCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 660  |
| RESULT 11   |  |   |        | Qy | 1278 | -----   | 1277 |
| LOCUS   | CQ771486                               | 1769 bp   | DNA    | Db | 661  | CCCCCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 720  |
| DEFINITION  | Sequence 1 from Patent WO2004009803.   |   | Linear | Qy | 1315 | GTGGAGAGCGGAGACACCAAGCTTGGGCGGCTGGGCGGCTGGGCAAGTCAAGCTT     | 1374 |
| ACCESSION   | CQ771486                               |   |        | Db | 721  | GTGGAGAGCGGAGACACCAAGCTTGGGCGGCTGGGCGGCTGGGCAAGTCAAGCTT     | 780  |
| VERSION   | CQ771486.1                             | GI:45125533   |        | Qy | 1375 | GGAGCACAAGCTCTGTGGGAGATCCCTGCTCTCCGAGGAGCTGGGCTGACAGCCGCCAC | 1434 |
| KEYWORDS  |  |   |        | Db | 781  | GGAGCACAAGCTCTGTGGGAGATCCCTGCTCTCCGAGGAGCTGGGCTGACAGCCGCCAC | 840  |
| SOURCE  |  |   |        | Qy | 1435 | TGCTTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG     | 1494 |
| ORGANISM  | Homo sapiens (human)                   |   |        | Db | 841  | TGCTTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG     | 900  |
| REFERENCE   | 1                                      |   |        | Qy | 1495 | CAGGCTCTTCCCAAGCTCTGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 1554 |
| AUTHORS   | Xiao, Y.                               |   |        | Db | 901  | CAGGCTCTTCCCAAGCTCTGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 960  |
| TITLE   | Regulation of human hepsin             |   |        | Qy | 1555 | CTTCCCTTTTGGGAGCCCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 1614 |
| JOURNAL   | Patent: WO 2004009803-A 1 29-JAN-2004; |   |        | Db | 961  | CTTCCCTTTTGGGAGCCCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG      | 1020 |
| FEATURES  | Bayer Healthcare AG (DE)               |   |        | Qy | 1615 | TTCAGTCCCTGCCCCCTCAAGAAATACATCAGCTGTGTGTGTGTGTGTGTGTGTGT    | 1674 |
| SOURCE  | Location/Qualifiers                    |   |        | Db | 1021 | TTCAGTCCCTGCCCCCTCAAGAAATACATCAGCTGTGTGTGTGTGTGTGTGTGTGT    | 1080 |
| ORIGIN  |  |   |        | Qy | 1675 | GCCCTGT     | 1734 |
| Query Match   | 59.5%; Score 1405; DB 6; Length 1769;  |   |        | Db | 1081 | GCCCTGT     | 1140 |
| Best Local Similarity   | 90.0%; Pred. No. 3.5e-259;             |   |        | Qy | 1735 | CAACAGGCGGGGATCTCCAGAGGCTGAGTCCCATATACAGATATATGTCTGCAAT     | 1794 |
| Matches 1592; Conservative 0; Mismatches 0; Indels 177; Gaps 1; |  |   |        | Db | 1141 | CAACAGGCGGGGATCTCCAGAGGCTGAGTCCCATATACAGATATATGTCTGCAAT     | 1200 |
| Qy  | 772                                    | GTCCACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG    | 831    | Qy | 1795 | GGGCTGACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGTGTGTGTGTGT  | 1854 |
| Db  | 1                                      | GTCCACCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG    | 60     | Db | 1201 | GGGCTGACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGTGTGTGTGTGT  | 1260 |
| Qy  | 832                                    | CAGAAGAGGAGTGGCCGAGCTGTGCAATGCTTCAGAACCCAGGAGGAGGAGGAGGAGG  | 891    | Qy | 1855 | GGTGGCAATTGATGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG   | 1914 |
| Db  | 61                                     | CAGAAGAGGAGTGGCCGAGCTGTGCAATGCTTCAGAACCCAGGAGGAGGAGGAGGAGG  | 120    | Db | 1261 | GGTGGCAATTGATGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG   | 1320 |
| Qy  | 892                                    | GGGGGAGCCCTGTCTTCTTGAAGCAATCGGGGCGGAGATCTGGGCGCATTTGTGGCTTT | 951    | Qy | 1915 | TCTGGAGCGCAGGTTGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  | 1974 |
| Db  | 121                                    | GGGGGAGCCCTGTCTTCTTGAAGCAATCGGGGCGGAGATCTGGGCGCATTTGTGGCTTT | 180    | Db | 1321 | TCTGGAGCGCAGGTTGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  | 1380 |
| Qy  | 952                                    | CTCTCAGAGATGACAGAGAGCCGCTGTACCCAGTGAAGTCAAGTCTTGCGGAGCTCGG  | 1011   | Qy | 1975 | GCCAGAGAGCAGGCGCTTACCAAAAGTCAAGTCACTTCGCGGAGTGGATCTTTCAGAGC | 2034 |
| Db  | 181                                    | CTCTCAGAGATGACAGAGAGCCGCTGTACCCAGTGAAGTCAAGTCTTGCGGAGCTCGG  | 240    | Db | 1381 | GCCAGAGAGCAGGCGCTTACCAAAAGTCAAGTCACTTCGCGGAGTGGATCTTTCAGAGC | 1440 |
| Qy  | 1012                                   | CTCATGTCTTTTGAACAAGCGAAAGGAGCTGGCGGCTGTGTCTCTCGCGCTTCAAC    | 1071   | Qy | 2035 | ATTAAGACTCACTCCAGAGCAGAGGAGTGTGATCCAGGCTTGAACCGGCTTCTCGC    | 2094 |
| Db  | 241                                    | CTCATGTCTTTTGAACAAGCGAAAGGAGCTGGCGGCTGTGTCTCTCGCGCTTCAAC    | 300    | Db | 1441 | ATTAAGACTCACTCCAGAGCAGAGGAGTGTGATCCAGGCTTGAACCGGCTTCTCGC    | 1500 |
| Qy  | 1072                                   | GGCGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG     | 1131   | Qy | 2095 | TGGCAGCTTCCAGAGGCGGAGGAGTATCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT  | 2154 |
| Db  | 301                                    | GGCGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG     | 360    | Db | 1501 | TGGCAGCTTCCAGAGGCGGAGGAGTATCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT  | 1560 |
| Qy  | 1132                                   | GAGGTGAGCTGCGAAACGCGCGGCGCAATGGACAGTCTGCTGTGTGTGTGTGTGTGT   | 1191   |    |      |   |      |
| Db  | 361                                    | GAGGTGAGCTGCGAAACGCGCGGCGCAATGGACAGTCTGCTGTGTGTGTGTGTGTGT   | 420    |    |      |   |      |
| Qy  | 1192                                   | GGAGAGTGTCCCAACATCCAGAGGCTGTGAGGTCACTTCGTGTGTGTGTGTGTGTGT   | 1251   |    |      |   |      |



QY 2155 GGAGTTTCTTCTTGGGCCCCGNCACAGGTCCAAAGACACCTCCCTCCAGGCTCT 2214  
DB 1561 GGAGCTTTTCTTCTTGGGCCCCGNCACAGGTCCAAAGACACCTCCCTCCAGGCTCT 1620  
QY 2215 CTCTTCCACAGTGGGCGGCCCACTCAGCCCCGAGACCAACCTCACTCCCTGAGACC 2274  
DB 1621 CTCTTCCACAGTGGGCGGCCCACTCAGCCCCGAGACCAACCTCACTCCCTGAGACC 1680  
QY 2275 CCATGTAAATATTTGTTCTGCTGTGCGGACTCTCTGTCTAAGTCCCTGATGATGAGATG 2334  
DB 1681 CCATGTAAATATTTGTTCTGCTGTGCGGACTCTCTGTCTAAGTCCCTGATGATGAGATG 1740  
QY 2335 CTCTTAAATATTAAGATGTTTGATT 2363  
DB 1741 CTCTTAAATATTAAGATGTTTGATT 1769

RESULT 12  
AR255885 1615 bp DNA linear PAT 20-DEC-2002  
LOCUS AR255885 Sequence 1 from patent US 6482630.  
DEFINITION AR255885  
ACCESSION AR255885  
VERSION AR255885.1 GI:27305128  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1..1615)  
AUTHORS Gan, W., Ye, J., DiFrancesco, V. and Beasley, E.M.  
TITLE Isolated human protease proteins, and uses thereof  
JOURNAL human protease proteins, and uses thereof  
FEATURES  
source 1..1615  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 53.7%; Score 1269; DB 6; Length 1615;  
Best Local Similarity 91.1%; Pred. No. 4.2e-233;  
Matches 1411; Conservative 0; Mismatches 15; Indels 123; Gaps 1;

QY 815 GAGCAGTGAATGCGCGCAAGAGAGGTGGCCGAGCTGTGCTGCTTCCAGACCA 874  
DB 167 GTGGCCCCAGCATGTGTGTGTGTCAGGTGGCCGAGCTGTGCTTCCAGACCA 226  
QY 875 AGGTGGCAGCTCTCAGTGGGAGACCCGTACTTCTGACAGCCATGCGGGCGGCTCT 934  
DB 227 AGGTGGCAGCTCTCAGTGGGAGACCCGTACTTCTGACAGCCATGCGGGCGGCTCT 286  
QY 935 GGGCCATTGTGGCTGTCTCTCAGAGTGAACAGAGCCGCTGTACCAAGTGAAGTCA 994  
DB 287 GGGCCATTGTGGCTGTCTCTCAGAGTGAACAGAGCCGCTGTACCAAGTGAAGTCA 346  
QY 995 GCTCTGCGGAGCGCTGGCTCATGTCTTTGACAGACGGAAGGACGTGGCGGCTGCT 1054  
DB 347 GCTCTGCGGAGCGCTGGCTCATGTCTTTGACAGACGGAAGGACGTGGCGGCTGCT 406  
QY 1055 GCTCTGCGGAGCGCTGGCTCATGTCTTTGACAGACGGAAGGACGTGGCGGCTTCTCA 1114  
DB 407 GCTCTGCGGAGCGCTGGCTCATGTCTTTGACAGACGGAAGGACGTGGCGGCTTCTCA 466  
QY 1115 GGGCAGTGAACCACTCCAGTGAAGTGAACGCGGCGGCAATGGAAGTCAAGTCA 1174  
DB 467 GGGCAGTGAACCACTCCAGTGAAGTGAACGCGGCGGCAATGGAAGTCAAGTCA 466  
QY 1175 TCTTCTGTGTGAGAGAGGAGAGGCTGCCACACCAAGGCTGTGAGAGTCACTCCG 1234  
DB 467 TCTTCTGTGTGAGAGAGGAGAGGCTGCCACACCAAGGCTGTGAGAGTCACTCCG 466  
QY 1235 TGTGTGATTTGCCCGAGAGCGGTTTCTTGGCGGCACTGTGCAAGACTGTGGCGGAG 1294  
DB 1235 TGTGTGATTTGCCCGAGAGCGGTTTCTTGGCGGCACTGTGCAAGACTGTGGCGGAG 1294

DB 467 ---GNGATGCCCCAGAGCGGTTTCTTGGCGGCACTGTGCAAGACTGTGGCGGAG 523  
QY 1295 AGCTGCCGTGTGAGACCGCATGTGTGAGAGCGGAGACACCACTTGGGCGGCTGTGGC 1354  
DB 524 AGCTGCCGTGTGAGACCGCATGTGTGAGAGCGGAGACACCACTTGGGCGGCTGTGGC 583  
QY 1355 AAGTCAGCTTGGCTATGATGATGAGACACCTCTGTGGGAGATCCCTGCTCCGAGGACT 1414  
DB 584 AAGTCAGCTTGGCTATGATGATGAGACACCTCTGTGGGAGATCCCTGCTCCGAGGACT 643  
QY 1415 GGGTGTGACAGCGGCCCACTCTTCCGAGAGGAAACCGGCTCTGTCCGATGAGAG 1474  
DB 644 GGGTGTGACAGCGGCCCACTCTTCCGAGAGGAAACCGGCTCTGTCCGATGAGAG 703  
QY 1475 TGTGTGCGGAGTGGCTGTGGCCCAAGGCTCTCCACAGGTCTGACGTGGGAGTGA 1534  
DB 704 TGTGTGCGGAGTGGCTGTGGCCCAAGGCTCTCCACAGGTCTGACGTGGGAGTGA 763  
QY 1535 TGTGTGACAGCGGCCCACTCTTCCGAGAGGAAACCGGCTCTGTCCGATGAGAG 1594  
DB 764 TGTGTGACAGCGGCCCACTCTTCCGAGAGGAAACCGGCTCTGTCCGATGAGAG 823  
QY 1595 ATATTGCGCTGTGTACCTCTCAGTCCCTGCGCTCAGAGAAATACAGCTGTGT 1654  
DB 824 ATATTGCGCTGTGTACCTCTCAGTCCCTGCGCTCAGAGAAATACAGCTGTGT 883  
QY 1655 GCTTCCAGTGTGGCGGCGGAGCCCTGTGTGAGAGAAATGTGATCCGTGACGGCTGG 1714  
DB 884 GCTTCCAGTGTGGCGGCGGAGCCCTGTGTGAGAGAAATGTGATCCGTGACGGCTGG 943  
QY 1715 GCAACAGCGAGTACTATGAGGCAAGGCGGAGGATCTCAGAGAGCTGATCCCATTA 1774  
DB 944 GCAACAGCGAGTACTATGAGGCAAGGCGGAGGATCTCAGAGAGCTGATCCCATTA 1003  
QY 1775 TCAGCAATGATGTGTGCAATGAGCGCTGATCTTATGAGAAACAGATCAAGCCAAATGT 1834  
DB 1004 TCAGCAATGATGTGTGCAATGAGCGCTGATCTTATGAGAAACAGATCAAGCCAAATGT 1063  
QY 1835 TGTGTGCTGTGATCCCGAGGAGTGGCATTTGATGCTTGCAGGCGAGACAGCGGTGCT 1894  
DB 1064 TGTGTGCTGTGATCCCGAGGAGTGGCATTTGATGCTTGCAGGCGAGACAGCGGTGCT 1123  
QY 1895 TGTGTGAGAGAGAGATCTCTGGAGCGGACGATGAGGCGGCTGTGTGAGATG 1954  
DB 1124 TGTGTGAGAGAGAGATCTCTGGAGCGGACGATGAGGCGGCTGTGTGAGATG 1183  
QY 1955 GGGCAGTGTGTGCTGTGCTGTGCGGAGAGCGGCTGTGATCAACCAAGTCAAGTCTTC 2014  
DB 1184 GGGCAGTGTGTGCTGTGCTGTGCGGAGAGCGGCTGTGATCAACCAAGTCAAGTCTTC 1243  
QY 2015 GGAAGTGAATCTTCAAGGCAATAAAGACTCACTCCGAAGCCAGGCGGATGTTGAC 2074  
DB 1244 GGAAGTGAATCTTCAAGGCAATAAAGACTCACTCCGAAGCCAGGCGGATGTTGAC 1303  
QY 2075 TGTGACGGTGTGCTGTGCGGAGCGCTTCCAGGCGGAGGATGATCCGCTGTGAG 2134  
DB 1304 TGTGACGGTGTGCTGTGCGGAGCGCTTCCAGGCGGAGGATGATCCGCTGTGAG 1363  
QY 2135 TCAAGCTGTGGGCGAGAGATGAGAGTCTTCTTGTGGCCCGGCTCCAGAGTCCAAAG 2194  
DB 1364 TCAAGCTGTGGGCGAGAGATGAGAGTCTTCTTGTGGCCCGGCTCCAGAGTCCAAAG 1423  
QY 2195 CACCTTCTTCCAGAGTCTCTTCTTCAAGTGGGCGGCGGCTCCAGAGTCCAAAG 2254  
DB 1424 CACCTTCTTCCAGAGTCTCTTCTTCAAGTGGGCGGCGGCTCCAGAGTCCAAAG 1483  
QY 2255 CAACCTACCTCTGAGCGGCGGCAATGTAATATGTTCTGAGTGTGGAGCTCGTGTAG 2314  
DB 1484 CAACCTACCTCTGAGCGGCGGCAATGTAATATGTTCTGAGTGTGGAGCTCGTGTAG 1543  
QY 2315 GTGCCCTGTATGATGAGTCTCTTAAATATTAAGATGTTTGATT 2363  
DB 1544 GTGCCCTGTATGATGAGTCTCTTAAATATTAAGATGTTTGATT 1592

RESULT 13  
LOCUS HSHPEPL 1199 bp mRNA linear PRI 22-SEP-1995  
DEFINITION H.sapiens liver mRNA for serine protease hepsin.  
ACCESSION X07002  
KEYWORDS X07002.1 GI:32065  
SOURCE hepsin; membrane protein; serine protease; zymogen.  
ORGANISM Homo sapiens (human)  
REFERENCE 1 (bases 1 to 1199)  
AUTHORS Leytus,S.P., Loeb,K.R., Hagen,F.S., Kurachi,K. and Davie,E.W.  
TITLE A novel trypsin-like serine protease (hepsin) with a putative  
JOURNAL transmembrane domain expressed by human liver and hepatoma cells  
MEDLINE Biochemistry 27 (3), 1067-1074 (1988)  
PUBMED 88209431  
COMMENT 2835076  
FEATURES  
source see X07732 for hepatoma hepsin full length cDNA.  
Location/Qualifiers  
1..1199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HUM 1250"  
/issue\_type="liver"  
/clone\_id="pBR322"  
1..1199  
/codon\_start=1  
/product="hepsin"  
/protein\_id="CAA50058.1"  
/db\_xref="GI:899286"  
/db\_xref="UniProt/Swiss-Prot:P05981"  
/translation="TSGPFCVDEGRLLPHOTRLLELVISVCCPRPRFLAATCODGRRK  
LPVDRIVIGARDTSLGRMPQVSLRDANHLGSLSDWVLTAAHFFPERNRVLSRW  
RVFAGVAQASPHQLQSVQAVVTHGSLFPRDPSSENSDIALVLSPLPTERT  
QPVCLPAAGALVDGKI CTYGMGNTQYGGQAGVLEQARVPIISNDVCAGADYQVQ  
IKPMFCAGYBEGGIDACQDSGSPFYCEDSISRTPRWRICGIVSMGTGALAKPKGV  
YTKVSDPREWIFQAIKTHSEASGMVTL"  
1..147  
148..902  
1199  
sig\_peptide  
mat\_peptide  
polya\_site  
ORIGIN  
Query Match 50.7%; Score 1199; DB 9; Length 1199;  
Best Local Similarity 100.0%; Pred. No. 1.2e-219; Mismatches 0; Indels 0; Gaps 0;  
Matches 1199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1165 ACGTCGAGCTTCTTCTGTGTGAGCAGAGGAGGCTGCCACACCCAGAGGCTCTGAG 1224  
DB 1 ACGTCGAGCTTCTTCTGTGTGAGCAGAGGAGGCTGCCACACCCAGAGGCTCTGAG 60  
QY 1225 GTGATCTCCGTGTGTGATGTGCGCCAGAGGCGTTTCTTGGCGGCATCTGCGAAGACTGT 1284  
DB 61 GTGATCTCCGTGTGTGATGTGCGCCAGAGGCGTTTCTTGGCGGCATCTGCGAAGACTGT 120  
QY 1285 GGCAGGAGAACTGCGCGTGGAGCCGATCGTGGAGAGCGCGGACACAGCTTGAGCGG 1344  
DB 121 GGCAGGAGAACTGCGCGTGGAGCCGATCGTGGAGAGCGCGGACACAGCTTGAGCGG 180  
QY 1345 TGGCGGTGGCAAGTCAGCTTGGTATGATGAGACACACTCTGTGGGAGATCCCTGCTC 1404  
DB 181 TGGCGGTGGCAAGTCAGCTTGGTATGATGAGACACACTCTGTGGGAGATCCCTGCTC 240  
QY 1405 TCCGGGAGATGGGTGTGACGACGCGCCACATGCTTCCCGAGCGGAAACCGGGTCTGTCC 1464  
DB 241 TCCGGGAGATGGGTGTGACGACGCGCCACATGCTTCCCGAGCGGAAACCGGGTCTGTCC 300  
QY 1465 CGATGCGAGTGTGTCGCGGTGCGTGGGCCAGGCTCTCCCAAGGTCTGCACTGGGG 1524

DB 301 CGATGCGAGTGTGTCGCGGTGCGTGGGCCAGGCTCTCCCAAGGTCTGCACTGGGG 360  
QY 1525 GTGAGGCTGTGTGCTACACAGGAGGCTATCTTCCCTTTGGGACCCCAACAGGAGAG 1584  
DB 361 GTGAGGCTGTGTGCTACACAGGAGGCTATCTTCCCTTTGGGACCCCAACAGGAGAG 420  
QY 1585 AACAGCAAGATATGCGCTGTGTGACCTGTCCAGTCCCTGCGCCCTCAGAGATATCATC 1644  
DB 421 AACAGCAAGATATGCGCTGTGTGACCTGTCCAGTCCCTGCGCCCTCAGAGATATCATC 480  
QY 1645 CAGCTGTGTGCTTCCAGCTGTGCGCGGACAGGCTGTGTGATGGCAAGTCTGATCCGTG 1704  
DB 481 CAGCTGTGTGCTTCCAGCTGTGCGCGGACAGGCTGTGTGATGGCAAGTCTGATCCGTG 540  
QY 1705 ACGGCTGTGGGCAACAGCGCTACTATGAGCAACAGGCGGGGTATCTCCAGAGGCTCGA 1764  
DB 541 ACGGCTGTGGGCAACAGCGCTACTATGAGCAACAGGCGGGGTATCTCCAGAGGCTCGA 600  
QY 1765 GTCCCATATACAGATGATGTGTGCAATGGGCTGACTTCTATGGAACAGATCAAG 1824  
DB 601 GTCCCATATACAGATGATGTGTGCAATGGGCTGACTTCTATGGAACAGATCAAG 660  
QY 1825 CCCAAGATGTGTGTGTGTGCTGCTACCCCGAGGCTGCAATTGATGCTGCGCAGGCGA 1884  
DB 661 CCCAAGATGTGTGTGTGTGCTGCTACCCCGAGGCTGCAATTGATGCTGCGCAGGCGA 720  
QY 1885 GGTGTCTCTTGTGTGTGTGAGACAGCATCTTCTGCAACGCGCAGTGTGGGCTGTGGC 1944  
DB 721 GGTGTCTCTTGTGTGTGTGAGACAGCATCTTCTGCAACGCGCAGTGTGGGCTGTGGC 780  
QY 1945 ATTGTGATGGGCACTGTGCTGTGCGCTGCGCCAGAGGCGAGGCTCTACCAAGATC 2004  
DB 781 ATTGTGATGGGCACTGTGCTGTGCGCTGCGCCAGAGGCGAGGCTCTACCAAGATC 840  
QY 2005 AGTGACTTCCGGAGATGTGATCTTCCAGGCGCATTAAGACTCACTCCGAAGCGAGT 2064  
DB 841 AGTGACTTCCGGAGATGTGATCTTCCAGGCGCATTAAGACTCACTCCGAAGCGAGT 900  
QY 2065 GTGACCAAGCTGTGACCGGTGTGCTTCTGCTGCGCAGGCTTCCAGGCGCGAGTGTATCC 2124  
DB 901 GTGACCAAGCTGTGACCGGTGTGCTTCTGCTGCGCAGGCTTCCAGGCGCGAGTGTATCC 960  
QY 2125 GGTGTGAGATTCACGCTGTGGGCGAGGATGAGAGGTTTTCTTGGGCGCGGTCCACA 2184  
DB 961 GGTGTGAGATTCACGCTGTGGGCGAGGATGAGAGGTTTTCTTGGGCGCGGTCCACA 1020  
QY 2185 GGTCAAGAGACACCTTCCCTCAAGGCTCTCTTCCACAGTGGCGGCTCACTAGCC 2244  
DB 1021 GGTCAAGAGACACCTTCCCTCAAGGCTCTCTTCCACAGTGGCGGCTCACTAGCC 1080  
QY 2245 CAGAGCAACCACTCACTCTCTGACCCCTCATATATGTTCTGTGTGTGGAG 2304  
DB 1081 CAGAGCAACCACTCACTCTCTGACCCCTCATATATGTTCTGTGTGTGGAG 1140  
QY 2305 TCCGTGTAGTGTGCGCTGTGATGGATGCTCTTAAATATTAAGATGTTTGAAT 2363  
DB 1141 TCCGTGTAGTGTGCGCTGTGATGGATGCTCTTAAATATTAAGATGTTTGAAT 1199

RESULT 14  
LOCUS RNHEPA 1739 bp mRNA linear ROD 19-JUL-1993  
DEFINITION R.norvegicus mRNA for hepsin.  
ACCESSION X70900  
VERSION X70900.1 GI:57928  
KEYWORDS hepsin; protease; serine protease.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
REFERENCE 1  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

AUTHORS Farley, D., Raymond, F. and Nick, H.  
 TITLE Cloning and sequence analysis of rat hepsin, a cell surface serine  
 protease  
 JOURNAL Biochim. Acta 1173 (3), 350-352 (1993)  
 MEDLINE 93305733  
 PUBMED 8318546  
 REFERENCES 2 (bases 1 to 1739)  
 AUTHORS Farley, D.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1993) D.L. Farley, Ciba Geigy, K125.117, 4002  
 BASEL, SWITZERLAND

FEATURES  
 source  
 1..1739  
 location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague Dawley"  
 /db\_xref="taxon:10116"  
 /sex="male"  
 /tissue\_type="liver"  
 /dev\_stage="7-8 week"  
 1..1739  
 /gene="hepsin"  
 1..1739  
 /evidence="experimental"  
 185..1435  
 /gene="hepsin"  
 /codon\_start=1  
 /product="hepsin"  
 /protein\_id="CAA50256.1"  
 /db\_xref="GI:57929"  
 /db\_xref="GDB:100551"  
 /db\_xref="UniProt/Swiss-Prot:Q05511"  
 /translation="MAKGGRTAPCCSRPKVAALIVGLFLTGIGASMAIVTLR  
 SDOELVQVQLSPDSRLVADKTEGTRILCSRSNRAVGLCEBGFRLAHSE  
 LDVRTAGDSIGSFCEVDEGGLPLAORLLDIVISDCPRGRPLATCODGRRLPVD  
 RIVGGDSILGRMPQVSLRYDGLHLCGSLISGDMWLTALCFPERRLVSRPVA  
 GAVARTSPHAYVQLGVQAVIYHGILPPEDPTIDENSNDIALVHSSSLPLEYIOPVC  
 LPAAGQALVDKVCITGWTQTFQDQAVVLEARVPTISNEVNSDFGNDIKPK  
 MFCAGYPEGIDACCGSGGHPCEDRISGTSRMLCGIVSMGTGCAIARPKGVYTV  
 IDPREWIFQAIKTHSEATGMVTP"  
 185..667  
 /gene="hepsin"  
 /gene="hepsin"  
 /gene="hepsin"  
 /product="hepsin"  
 /product="hepsin"

ORIGIN  
 Query Match 43.3%; Score 1023.6; DB 10; Length 1739;  
 Best Local Similarity 79.3%; Pred. No. 4.6e-186;  
 Matches 1314; Conservative 0; Mismatches 309; Indels 35; Gaps 7;

Db 388 AGAGGAAACGTGAGAGCTGCTGCTCTCTCAACGCTCCAAAGCCAGAGGATGAGAGGCTCG 447  
 1092 CTGCAGAGAGATGGGCTTCTCTCAGAGGACCTGACCACTCCAGAGTGAAGTGCAGAGCGC 1151  
 448 CTGTGAGAGATGGGCTTCTCAGAGGCTCTGCGGACCTGACAGGCTGATGTGCGAAGCGC 507  
 1152 GGGCGCCAAATGGCAAGTGGGCTTCTTGTGTGTGAGCAAGGGAGAGGCTGCCCAACCCA 1211  
 508 GGGGCGCCAAAGGACATCGGGCTTCTTGTGTGTGAGCAAGGGAGAGGCTGCCCTGACCTCA 567  
 1212 GAGGCTGCTGAGAGGATCTCCGTGTGTGATGTCGCCAGAGAGCGGCTTCTTGGCCGCCAT 1271  
 568 GCGGTGCTGAGATGTCACTCTGTATGTGAGCATGTCTCTAGAGCCGATTCCTGACTGACCA 627  
 1272 CTGCCAAGACTGTGTGGCCGACAGAAAGCTGCCGTGACCGCATCTGTGGAGGCCGGAGAC 1331  
 628 CTGCCAAGACTGTGTGGCCGACAGAAAGCTGCCGTGATGCAATGTGTGGGGGACAGACAG 687  
 1332 CAGCTTGGGCGGCTGGCCGTCAGAAAGTCAAGCTTCTGCTATGATGAGACACACTTGTGG 1391  
 688 CAGCCTGGGAAGATGGCCATGGCAGGTCAAGCTCGATTATGATGGAAACCAACTTGTGG 747  
 1392 GGAATCCCTGCTCCGGGAGCTGGAGTGTGACAGCCGCCACTGCTCCCGAGCGGAA 1451  
 748 GGAATCCCTGCTGCTCCGGGAGCTGGAGTGTGACAGCCGCCACTGCTCCCGAGCGGAA 807  
 1452 CCGGGCTCTGTCCCGATGGCGAGTGTGGCGGAGTGGCCGAGCCGCTCTCCCGACGG 1511  
 808 CCGGGCTCTGTCTCGGTGGCGAGTATTGTGTGTGTGATGCTCCGGACCTCACTCATGCG 867  
 1512 TCTGCACTGGGGGTGCAAGGCTGTGTGTACCAAGGGGGCTATCTTCCCTTGGGAGCC 1571  
 868 GGTCACTGGGGGTTCAGAGCTGTGATCAATGAGGGGCTATCTTCCCTTGGGAGCC 927  
 1572 CAACAGGAGAGAGAAACAGACGATGATGACCTGATCCACTCTCCAGTCCCTGCGCCCT 1631  
 928 TACTATGACGAAACAGCAATGACATGCTGCTGATCCACTCTCTAGCTCCCTGCTCT 987  
 1632 CACAGAAATACATTCAGCCTGTGTGCTCCAGCTGCGCGGACAGCCCTGTGTGATGGCAA 1691  
 988 CACAGAAATACATTCAGCCTGTGTGCTCCAGCTGCGCGGACAGCCCTGTGTGATGGCAA 1047  
 1692 GATCTGTACCTGACCGGCTGTGGGCAACAGCAATGATGATGATGATGATGATGATGATG 1751  
 1048 GGTCTGTACATGTACCGGCTGTGGGCAACAGCAATGATGATGATGATGATGATGATG 1107  
 1752 CCAGAGGCTGTGATCCCAATCAGCAATGATGATGATGATGATGATGATGATGATGATG 1811  
 1108 CCAAGAGGCTGTGATCCCAATCAGCAATGATGATGATGATGATGATGATGATGATGATG 1167  
 1812 AAACCAATCAAGCCCAAGATGTTCTGTGCTGTGATCCCAAGGAGTGCATGTGATGCTGT 1871  
 1168 GAATCAATCAAGCCCAAGATGTTCTGTGCTGTGATCCCAAGGAGTGCATGTGATGCTGT 1227  
 1872 CCAGGCGACAGCGGTGCTCTTGTGTGTGAGCAAGCATCTCTGTGAGCGCAAGTGTG 1931  
 1228 CCAGGCGACAGCGGTGCTCTTGTGTGTGAGCAAGCATCTCTGTGAGCGCAAGTGTG 1287  
 1932 GCGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1991  
 1288 GCGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1347  
 1992 CTACACCAAGTCAAGTCACTTCCGAGAGTGTGATCTTCCAGGCAATTAAGATCACTCGA 2051  
 1348 GTACACCAAGTCAAGTCACTTCCGAGAGTGTGATCTTCCAGGCAATTAAGATCACTCGA 1407  
 2052 AGCCAGCGCATGTGTACCAAGTCTGAC-----CGGTGCTTCTGCTGCG-6CA 2100  
 1408 AGCTACCGGATGTGTACCTGAGCCCTGACCCCGGCTCATGCTGCTGCGGCTGCTCA 1467  
 2101 GCGTCAAGGCGCAGAGTATC-----CCGGTGTGGAGATCAAGCTGCG 2145

|                            |   |   |                 |
|----------------------------|---|---|-----------------|
| Db                         | 1468  | GCATCCGAGAGTCAGAGTTGGCTCTGATGGCTCCACCCGACAGTGGAGGCTCCACACTGGG | 1527            |
| Qy                         | 2146  | CCGAGAGATGGACGTTTTCTTTCTTTGGGCCCGGTCCACAGGTCCAAGACACCTCCCTC   | 2205            |
| Db                         | 1528  | CCTCAACATGGAACGGTTTCTTGCTCGGATCCAGTCCATAGATCAAGG--ATGCTGGATC  | 1585            |
| Qy                         | 2206  | CAGGTCCTCTCTTCCACAGTGGCGGGCCCACTCAGCCCCGAGAACCAACCACTCAACC    | 2265            |
| Db                         | 1586  | CAAGGACCTCTCTTCCACAGTGGCGGGCCCACTCAATCCAGGAGCCATTGGCTCAACCT   | 1645            |
| Qy                         | 2266  | TCCTGACCCCCATGTAATATTTGTTCTGCTGTGGAGCTCTGTCTAGAGTCCCCGTAT     | 2325            |
| Db                         | 1646  | CCC--ACCCATGTAATATTACTCTGTCTCTGGG--GGCTGCTTTCCAGGCGCCCTT      | 1701            |
| Qy                         | 2326  | GATGGAGATGCTCTTTAATAATAAAGATGGTTTGATT                         | 2363            |
| Db                         | 1702  | GTGGGAGATGCTTTAATAATAAAGATGGTTTGATT                           | 1739            |
| RESULT 15                  |   |   |                 |
| LOCUS                      | AF030065  | 1781 bp   | mRNA            |
| DEFINITION                 | Mus musculus serine protease hepsin mRNA, complete cds.   |   | ROD 12-NOV-1997 |
| ACCESSION                  | AF030065  |   |                 |
| VERSION                    | AF030065.1  | GI:2606036  |                 |
| KEYWORDS                   |   |   |                 |
| SOURCE                     | Mus musculus (house mouse)  |   |                 |
| ORGANISM                   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.      |   |                 |
| REFERENCE                  | 1 (bases 1 to 1781)   |   |                 |
| AUTHORS                    | Vu,T.-K.H., Liu,R.W., Haaksma,C., Tomasek,J.J. and Howard,E.W.  |   |                 |
| TITLE                      | Identification and cloning of the membrane-associated serine protease, hepsin, from mouse preimplantation embryos                         |   |                 |
| JOURNAL                    | J. Biol. Chem. (1997) In press  |   |                 |
| REFERENCE                  | 2 (bases 1 to 1781)   |   |                 |
| AUTHORS                    | Vu,T.-K.H., Liu,R.W., Haaksma,C., Tomasek,J.J. and Howard,E.W.  |   |                 |
| TITLE                      | Direct Submision  |   |                 |
| JOURNAL                    | Submitted (14-OCT-1997) Pathology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd, Oklahoma City, OK 73104, USA |   |                 |
| FEATURES                   |   |   |                 |
| source                     | Location/Qualifiers   |   |                 |
|                            | 1..1781   |   |                 |
|                            | /organism="Mus musculus"  |   |                 |
|                            | /mol_type="mRNA"  |   |                 |
|                            | /db_xref="taxon:10090"  |   |                 |
|                            | /tissue_type="liver"  |   |                 |
|                            | 205..1455   |   |                 |
|                            | /note="serine protease"   |   |                 |
|                            | /codon_start=1  |   |                 |
|                            | /product="hepsin"   |   |                 |
|                            | /protein_id="AAB84221.1"  |   |                 |
|                            | /db_xref="GI:2606037"   |   |                 |
|                            | /translation="MAKEGRTAACCSRPVAAALIVGTLPLFTGIGASMAVITLLQ   |   |                 |
|                            | SDDEPLTVQVLSPEDSRLVLDITBETGIMWLCSSRNNAVAGLGEEMFPLALHSE  |   |                 |
|                            | LVDRTAANGTSGRFCVDEBGLPLAQRLLDVISVDCPEGRPLTICDCGKRKLPS   |   |                 |
|                            | RIVAGQDSISLRMPQVSLRVDGTHLCGSLISGDMVLTLAALCFEPRNRVLSRMTVA  |   |                 |
|                            | GAVARSPHAVOVAVIYHGGYLPFRPPTIDENNDNALVHLSSSLPLTAYIQVC  |   |                 |
|                            | LPAAGALVNGKVTCTVGMGNTOPVGQAMVLOEAVPIISNEVNSDPFVQNOIKPK  |   |                 |
|                            | MFCAIGREGSIDACQDSGSPFYCEDSIGSTRMRRLCGIVSNGTGALAKRKPVYTKV  |   |                 |
|                            | TDREWLFRKIKTHSEASGNVTP"   |   |                 |
| ORIGIN                     |   |   |                 |
| Query Match                | 43.3%; Score 1022.2; DB 10; Length 1781;  |   |                 |
| Best Local Similarity      | 79.2%; Pred. No. 8.6e-186;  |   |                 |
| Matches 1314; Conservative | 0; Mismatches 308; Indels 37; Gaps 7;   |   |                 |
| Qy                         | 732   | GGCTCCCCCAAGGCGCTCCCTCCGATCTCTCTCAAGGTCACACCTGGCCGAGAG        | 791             |
| Db                         | 111   | GACTTAACCCCAACCTGACCATCTCCGGGAAACCCAGGTTCCGCCCAACGAG          | 170             |
| Qy                         | 792   | GTCAGCCAGGATCATTAACAAGAGCACTGGCGCAGAGAGAGGTTGGCCGAGC          | 851             |

```
OY 1932 GCGGCTGTTGGCATTGTAAGTTGGGGGCACTGGCTGTGCTCCCTGGCCAGAAAGCCAGGGCT 1991
    |||||
Db 1308 GCGGCTATGTGGCATTGTAAAGTGGGGTACCGGCTGTGCTTTGGCTCGAAGCCAGAGT 1367
    |||||
OY 1992 CTACACCAAAAGTCAGTGACTTCCGGAGTGAATCTTCCAGGCCATAAAGACTCACTCCGA 2051
    |||||
Db 1368 GTACACCAAAAGTCAGTGACTTCCGGAGTGAATCTTCCAGGCCATAAAGACTCACTCCGA 1427
    |||||
OY 2052 AGCCAGCGGCATGTGACCCAGCTCTGA-CCGGTGGCTTCTGCTGGC----- 2099
    |||||
Db 1428 AGCCAGTGGCATGTGACTCAAGCCCTGATCCGCTCATCTGCTGCTCCGTTGCACT 1487
    |||||
OY 2100 AGCTCCAGGGGCCGAGGTGA------CCGGTGGTGGATCCAGCTGG 2144
    |||||
Db 1488 AGCATCCAGAGTCAGAGTTGGTCTGGTGGCTCCAGCCCAAGTGTAGGCTCCACACTGG 1547
    |||||
OY 2145 GCGGAGATGGGACGTTTTTCTTCTTGGGCCCGGTCCACAGGTCCAGAGACCTCCCT 2204
    |||||
Db 1548 GCTTCACATGGAATGTTTCTGCTCAGATCCAGTCCAGGGTCCAGG--ATGCTGGAT 1605
    |||||
OY 2205 CCAGGGTCTCTTCTTCCACAGTGGGGGCCCACTGAGCCCGAGACCAACCACTCACC 2264
    |||||
Db 1606 CCAAGGACTTCTCTTCCACAGTGGGCCGCCCACTCAATCCAGGGCCATGGGCTCACCC 1665
    |||||
OY 2265 CTCTGACCCCCCATGTAAATATGTCTGCTGTCTGGACTCTGTCTAGTGGCCCTGA 2324
    |||||
Db 1666 TCCC--ACCCCATGTAAATATATCTGTCTCTGCGGGGGCGCTTAAGGAGCCCT-- 1720
    |||||
OY 2325 TGATGGATGCTCTTAAATATAAAGATGTTTGATT 2363
    |||||
Db 1721 TGTGAGATGCTCTTTAAATATAAAGGTGGTTTGATT 1759
    |||||
```

Search completed: August 17, 2005, 07:27:19  
Job time : 16716 secs

**this Page blank (uspto)**